

Volatility Inference and Return Dependencies in Stochastic Volatility Models

Oliver Pfante ^{*1} and Nils Bertschinger ^{†1}

¹Systemic Risk Group, Frankfurt Institute for Advanced Studies

October 4, 2016

Abstract

Stochastic volatility models describe stock returns r_t as driven by an unobserved process capturing the random dynamics of volatility v_t . The present paper quantifies how much information about volatility v_t and future stock returns can be inferred from past returns in stochastic volatility models in terms of Shannon's mutual information.

1 Introduction

Many plaudits have been aptly used to describe Black and Scholes' [4] contribution to option pricing theory. However, especially after the 1987 crash, the geometric Brownian motion model and the Black-Scholes formula were unable to reproduce the option price data of real markets. This is not surprising, since the Black-Scholes model makes the strong assumption that log-returns of stocks are independently normally distributed with a volatility which is not only assumed to be known but also constant over time. Both assumptions are wrong: first, there are long range dependencies among returns; second, volatility is a hidden parameter which needs to be inferred from stock and option data, respectively, and is not constant at all but a highly volatile time-process. Among the most relevant statistical properties of these volatility stochastic processes, volatility seems to be responsible for the observed clustering in stock returns. That is, large returns are commonly followed by other large returns and similarly for small ones [7]. Another feature is that, in clear contrast with stock returns, which show negligible autocorrelations, squared stock return, which is essentially volatility, autocorrelation is still significant for time lags longer than one year [21, 20, 7, 18, 10, 17]. Additionally, there exists the so-called leverage effect, i.e., much shorter (few weeks) negative cross-correlation between current stock returns and future volatility [6, 7, 4, 5].

Inspired by these observations, stochastic volatility models have been developed which describes time-varying volatility as well as dependencies among stock returns. Here, we build on a range of generic stochastic volatility models empirically studied in [8] and the Exponential Ornstein-Uhlenbeck model. Stochastic volatility models describe volatility as a stochastic process in its own right. This process then couples to the stock price process, thereby trying to capture many prominent statistical properties of stock returns, temporal clustering, leveraging, volatility autocorrelation, as well as the volatility smile observed in option prices. In the present

^{*}pfante@fias.uni-frankfurt.de

[†]bertschinger@fias.uni-frankfurt.de

paper we pose two question on stochastic volatility models: first, how reliably can the hidden volatility be inferred from observable stock-return data; second, how dependent are subsequent returns in these models? Since, in stochastic volatility models stock prices as well as the time varying volatility are both modelled as random variables, Shannon’s information theory [24] provides an ideal frame to make this question precise. In particular, in this paper we address the question of how much information observed stock returns provide about the hidden volatility and future returns.

To this end, we first introduce basic information theory in section 2.1. In particular, we recall the definition of mutual information, a general measure of statistical dependence and provide an intuitive interpretation in the context of volatility estimation. In the subsequent subsection we recall the basics on existence and uniqueness of stationary solutions of the Fokker-Planck equation. Furthermore, we introduce the logarithmic Sobolev inequality which is vastly applied in the present paper. This inequality has played a major role in recent mathematics in estimating convergence rates of arbitrary time-dependent solutions of the Fokker-Planck equation towards the stationary one (if it exists) [19]. In section 3 we explain how this allows to quantify the amount of information obtained in stochastic volatility models and derive analytic upper bounds for a wide class of single factor stochastic volatility models in section 4. Using realistic parameter values fitted from the literature, we find that stock returns provide in general only limited information about the volatility and are fairly independent in stochastic volatility models on any time-scale. Our calculations reveal that this is not a data issue, but instead arises for quite fundamental information theoretic reasons. This not only implies that volatility estimates from stock data are inherently imprecise, but, as we believe, also has severe implications for volatility predictions and sheds doubt on the standard practice of comparing models based on their forecasting error [14]. The inability of stochastic volatility models to impose strong return dependencies suggests that their less complex single factor Jump-Diffusion models [12] and Lévy models [23], respectively, provide a reasonable alternative. Even though returns are independent in these models, they allow for large tails and lead to analytically tractable option pricing models that can generate implied volatility smiles. We conclude with a short comment on multi-factor stochastic volatility models.

2 Prerequisites

This section introduces basic elements of information theory and results on the Fokker Planck-equation.

2.1 Information Theory

The *differential entropy* of a tuple $\mathbf{X} = (X_1, X_2, \dots, X_n)$ of random variables with joint density $f(\mathbf{x}) = f(x_1, x_2, \dots, x_n)$ is defined as [9]

$$h(\mathbf{X}) = - \int f \log f$$

which might be $-\infty$. Here, and in the sequel, we generally drop the $d\mathbf{x}$ -notation in integrals w.r.t. the Lebesgue measure as long as no confusions occur and shall write dp if we integrate w.r.t. a measure p .

Example 2.1. Let \mathbf{X} be a n dimensional and normally distributed with mean $\boldsymbol{\mu}$ and covariance

matrix A . We compute

$$\begin{aligned}
h(\mathbf{X}) &= - \int \frac{1}{\sqrt{2^n \pi^n \det(A)}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T A^{-1}(\mathbf{x}-\boldsymbol{\mu})} \log \left(\frac{1}{\sqrt{2^n \pi^n \det(A)}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T A^{-1}(\mathbf{x}-\boldsymbol{\mu})} \right) \\
&= \frac{1}{2} \log(2^n \pi^n \det(A)) + \int \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T A^{-1} (\mathbf{x} - \boldsymbol{\mu}) \frac{1}{\sqrt{2^n \pi^n \det(A)}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T A^{-1}(\mathbf{x}-\boldsymbol{\mu})} \\
&= \frac{1}{2} \log(2^n \pi^n \det(A)) + \frac{1}{2} \\
&= \frac{1}{2} \log(2^n \pi^n e \det(A)) .
\end{aligned}$$

Differential entropy may become negative as example 2.1 for sufficiently small $\det(A)$ proves. Even though it might become negative, the differential entropy, as the entropy of discrete random variables, can be interpreted as a measure of the average uncertainty in the random variable. As shown in [9], the entropy corresponds to the logarithm of the volume of typical outcomes, i.e., a random variable is more tightly concentrated the smaller its entropy.

The differential entropy behaves nicely under diffeomorphic coordinate changes $\phi : S_{\mathbf{X}} \rightarrow \mathbb{R}^n$ on the support $S_{\mathbf{X}} = \{\mathbf{x} = (x_1, x_2, \dots, x_n) : f(\mathbf{x}) > 0\}$ of the random variables $\mathbf{X} = (X_1, X_2, \dots, X_n)$. We obtain

$$h(\phi(\mathbf{X})) = h(\mathbf{X}) + \int f \log |\det J_\phi| \quad (2.1)$$

where J_ϕ is the Jacobian of ϕ . If the random variables \mathbf{X}, \mathbf{Y} have a joint density function $f(\mathbf{x}, \mathbf{y})$ and conditional density function $g(\mathbf{x}|\mathbf{y})$, respectively, we can define the *conditional entropy* $h(\mathbf{X}|\mathbf{Y})$ as

$$h(\mathbf{X}|\mathbf{Y}) = - \int f \log g .$$

Since in general $f(\mathbf{x}|\mathbf{y}) = f(\mathbf{x}, \mathbf{y})/f(\mathbf{y})$, we can also write

$$h(\mathbf{X}|\mathbf{Y}) = h(\mathbf{X}, \mathbf{Y}) - h(\mathbf{Y}) .$$

But we must be careful if any of the differential entropies are infinite. $h(\mathbf{X}|\mathbf{Y})$ is a measure of the average uncertainty of the random variable X conditional on the knowledge of another random variable Y .

The *relative entropy* (or *Kullback Leibler divergence*) $D(f||g)$ between two densities f and g is defined by

$$D(f||g) = \int f \log \frac{f}{g} . \quad (2.2)$$

Note that $D(f||g)$ is finite only if the support set $\{\mathbf{x} : f(\mathbf{x}) > 0\}$ of f is contained in the support of g (Motivated by continuity, we set $0 \log(0/0) = 0$). While $D(f||g)$ is in general not symmetric, it is often considered as a kind of distance between f and g . This is mainly due to its property that $D(f||g) \geq 0$ with equality if and only if $f = g$.

The *mutual information* $I(\mathbf{X} : \mathbf{Y})$ between two random variables \mathbf{X} and \mathbf{Y} with joint density $f(\mathbf{x}, \mathbf{y})$ and respective marginal densities $g(\mathbf{x})$ and $k(\mathbf{y})$ is defined as

$$\begin{aligned}
I(\mathbf{X} : \mathbf{Y}) &= h(\mathbf{X}) - h(\mathbf{X}|\mathbf{Y}) \\
&= h(\mathbf{Y}) - h(\mathbf{Y}|\mathbf{X}) \\
&= h(\mathbf{X}) + h(\mathbf{Y}) - h(\mathbf{X}, \mathbf{Y}) \\
&= \int f \log \frac{f}{gk} \\
&= D(f||gk) .
\end{aligned} \quad (2.3)$$

Again, we must be careful if any of the differential entropies are infinite and the mutual information might or might not diverge in this case.

From the definition it is clear that the mutual information is symmetric

$$I(\mathbf{X} : \mathbf{Y}) = I(\mathbf{Y} : \mathbf{X}),$$

non-negative

$$I(\mathbf{X} : \mathbf{Y}) \geq 0$$

and equal 0 if and only if \mathbf{X} is independent from \mathbf{Y} . Thus, mutual information can be considered as a general measure of statistical dependence as it detects any deviations from independence. Note, that in the case of independence, knowledge of \mathbf{X} does not reduce our uncertainty about \mathbf{Y} , i.e., \mathbf{X} provides no information about \mathbf{Y} , and vice versa.

Like mutual information, *conditional mutual information* $I(\mathbf{X} : \mathbf{Y} | \mathbf{Z})$ between three random variables \mathbf{X}, \mathbf{Y} and \mathbf{Z} can be written in terms of conditional entropies as

$$I(\mathbf{X} : \mathbf{Y} | \mathbf{Z}) = h(\mathbf{X} | \mathbf{Z}) - h(\mathbf{X} | \mathbf{Y}, \mathbf{Z})$$

assuming the differential entropies exist. This can be rewritten to show its relationship to mutual information

$$I(\mathbf{X} : \mathbf{Y} | \mathbf{Z}) = I(\mathbf{X} : \mathbf{Y}, \mathbf{Z}) - I(\mathbf{X} : \mathbf{Z})$$

usually rearranged as the *chain rule of mutual information* [9]

$$I(\mathbf{X} : \mathbf{Y}, \mathbf{Z}) = I(\mathbf{X} : \mathbf{Y} | \mathbf{Z}) + I(\mathbf{X} : \mathbf{Z}). \quad (2.4)$$

Definition 2.1. Suppose \mathbf{X}, \mathbf{Y} and \mathbf{Z} are three random variables. We call

$$\mathbf{X} \rightarrow \mathbf{Y} \rightarrow \mathbf{Z}$$

a *Markov Chain* also written as

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{Y}$$

if

$$I(\mathbf{X} : \mathbf{Z} | \mathbf{Y}) = 0,$$

i.e., the random variable \mathbf{X} is independent from \mathbf{Z} given \mathbf{Y} .

From the chain rule Eq. (2.4) we obtain the *Data Processing Inequality*

Corollary 2.1. If $\mathbf{X} \rightarrow \mathbf{Y} \rightarrow \mathbf{Z}$ is a Markov Chain, then

$$I(\mathbf{Z} : \mathbf{X}) \leq I(\mathbf{Y} : \mathbf{X})$$

with identity if and only if $\mathbf{X} \rightarrow \mathbf{Z} \rightarrow \mathbf{Y}$.

In contrast to differential entropy, mutual information and conditional mutual information are scaling invariant, that is, for three diffeomorphic maps ϕ_x, ϕ_y, ϕ_z we have [16]

$$\begin{aligned} I(\phi_x(\mathbf{X}) : \phi_y(\mathbf{Y})) &= I(\mathbf{X} : \mathbf{Y}) \\ I(\phi_x(\mathbf{X}) : \phi_y(\mathbf{Y}) | \phi_z(\mathbf{Z})) &= I(\mathbf{X} : \mathbf{Y} | \mathbf{Z}). \end{aligned} \quad (2.5)$$

Example 2.2. Let \mathbf{X} and \mathbf{Y} be jointly normally distributed with mean $\boldsymbol{\mu}$, marginal variances σ_X^2, σ_Y^2 and covariance σ_{XY} . We compute

$$\begin{aligned}
I(\mathbf{X} : \mathbf{Y}) &= h(\mathbf{X}) + h(\mathbf{Y}) - h(\mathbf{X}, \mathbf{Y}) \\
&= \frac{1}{2} \log(2\pi e \sigma_X^2) + \frac{1}{2} \log(2\pi e \sigma_Y^2) - \left[\frac{2}{2} \log(2\pi e) + \frac{1}{2} \log(\sigma_X^2 \sigma_Y^2 - \sigma_{XY}^2) \right] \\
&= \frac{1}{2} \log(\sigma_X^2 \sigma_Y^2) - \frac{1}{2} \log(\sigma_X^2 \sigma_Y^2 (1 - \rho_{XY}^2)) \\
&= -\frac{1}{2} \log(1 - \rho_{XY}^2)
\end{aligned}$$

where $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$ denotes the correlation coefficient between \mathbf{X} and \mathbf{Y} .

Similarly, we can ask how much information is required to estimate a random quantity \mathbf{X} up to some precision. Formally, let us assume that \mathbf{X} has a normal distribution with variance σ^2 . Now, having obtained a measurement the residual variance, i.e. the measurement error, is reduced to σ_M^2 . The benchmark is then provided by the following question: How much information do we need s.t. we feel confident of knowing \mathbf{X} up to m decimals? To this end, in the present setting we would require that $\sigma_M \approx \frac{1}{4} 10^{-m}$ motivated by the fact that \mathbf{X} would be distributed around the measured value \hat{x} with 95% confidence interval $\approx [\hat{x} - 2\sigma_M, \hat{x} + 2\sigma_M]$.

Correspondingly the required information is given by

$$h(\mathbf{X}) - h(\mathbf{X}_{\text{measured}}) = \frac{1}{2} \log(2\pi e \sigma^2) - \frac{1}{2} \log(2\pi e \sigma_M^2) = \log\left(\frac{\sigma^2}{\sigma_M^2}\right)$$

Thus, in the case of normal distributions, the required information is a non-linear function of the required reduction in uncertainty as measured by the ratio of standard deviations $\frac{\sigma}{\sigma_M} = 4 \cdot 10^m$.

2.2 Fokker-Planck Equation

We start with the $n + 1$ dimensional stochastic differential equation (SDE)

$$\begin{aligned}
dS_t &= \mu(\mathbf{v}_t) S_t dt + f(\mathbf{v}_t) S_t dW_t^0 \\
d\mathbf{v}_t &= \beta(\mathbf{v}_t) dt + \gamma(\mathbf{v}_t) d\mathbf{W}_t
\end{aligned} \tag{2.6}$$

S_t denotes the stock price at time t and \mathbf{v}_t is the volatility process driving the stock. Since we do not want to deal with regularity issues β, γ and f are smooth. Furthermore, we assume uniform ellipticity which implies that $\gamma(\mathbf{v}_t)$ is invertible for all $|\mathbf{v}_t|^2 > 0$. W_t^0 and $\mathbf{W}_t = (W_t^1, \dots, W_t^n)$ are 1 dimensional and n dimensional standard Wiener processes, respectively, which may be coupled, that is, we allow for $dW_t^0 d\mathbf{W}_t = (\rho_1, \dots, \rho_n) dt$ with $\rho_i \neq 0$ for $i = 1, \dots, n$. The second SDE of Eq. (2.6) corresponds to the linear Fokker-Planck equation

$$\partial_t p_t(\mathbf{v}) = -\nabla(\beta(\mathbf{v}) p_t(\mathbf{v})) + \frac{1}{2} \sum_{i,j=1}^n \partial_i \partial_j \left((\gamma(\mathbf{v}) \gamma^T(\mathbf{v}))_{ij} p_t(\mathbf{v}) \right) \tag{2.7}$$

for the probability density where $\partial_i = d/dv_i$ and $\nabla = (\partial_1, \dots, \partial_n)$. We call a solution p_t of the Fokker-Planck equation *stationary* if

$$\partial_t p_t(\mathbf{v}) = 0,$$

that is, p_t does no longer depend on t and we write $p_t = \rho$. We call the n dimensional volatility process a *gradient flow* perturbed by noise whose strength is 1 if the SDE reads

$$d\mathbf{v}_t = \sqrt{2} dW_t - \nabla V(\mathbf{v}_t) dt \tag{2.8}$$

with a smooth potential V . Eq. (2.8) corresponds to the linear Fokker-Planck equation

$$\partial_t p = \nabla \cdot (\nabla p + p \nabla V). \quad (2.9)$$

It is not possible to compute the time dependent solution for an arbitrary potential. However, e^{-V} is a solution of the ordinary differential equation

$$0 = \nabla \cdot (\nabla p + p \nabla V)$$

and if e^{-V} is integrable we can define the *Gibbs distribution*

$$\rho(x) = \frac{1}{Z} e^{-V} \quad (2.10)$$

where the normalization constant Z is the *partition function*

$$Z = \int e^{-V}.$$

In general, integrability of e^{-V} does not guarantee the existence of a stationary solution. We have a closer look on existence and uniqueness of such a stationary solution in the one dimensional case. We define the probability current

$$S = -\partial_v p - p \partial_v V$$

for a solution p of Eq. (2.9). For a stationary solution ρ the probability current must be constant. Thus, if it vanishes at some point v it must be zero for any v . If this is the case, we can immediately integrate the expression for the probability current which yields the expression Eq. (2.10). For a time dependent solution p of Eq. (2.9) the probability current may be written in the form

$$S = -e^{-V} \partial_v (e^V p)$$

In the stationary case, where S is constant, we thus have for arbitrary S

$$\rho = \frac{1}{Z} e^{-V} - S e^{-V} \int e^V$$

One of the integration constants is determined by the normalization

$$\int \rho = 1.$$

The other constant must be determined from the boundary conditions, so the problem arises which boundary conditions must be used. Feller [11] classified all boundary conditions in dimension one. Let denote with x_{\min} and x_{\max} the boundary points of the domain of the stationary distribution ρ (which includes also the cases $x_{\min} = -\infty$ and $x_{\max} = \infty$). The existence of such a stationary solution forces the probability current S to vanish at x_{\min} and x_{\max} and therefore $S = 0$ at any point, that is, x_{\min} and x_{\max} need to be *reflecting boundaries*. This imposes the *natural boundary conditions* $p(x_{\min}) = p(x_{\max}) = 0$ where p denotes a solution of the gradient Flow Eq. (2.9). In this case, any time dependent solution converges to the stationary solution Eq. (2.10) which proves uniqueness. We refer to chapter 5 in [22] for more details. We summarize these insights in

Lemma 2.2. *If Eq. (2.9) has a stationary solution ρ then*

1. $S \equiv 0$

2. e^{-V} is integrable
3. ρ is unique and concurs with the Gibbs distribution Eq. (2.10)
4. $\rho(x_{\min}) = \rho(x_{\max}) = 0$.

We return to the n dimensional case. If

$$\nabla^2 V \geq \lambda I_{n \times n} \quad \forall v \in \mathbb{R}^n \text{ and } \lambda > 0, \quad (2.11)$$

that is, the Hessian of V is uniformly bounded away from zero by a positive real λ , from [2] follows that $\rho = e^{-V - \log(Z)}$ fulfils the logarithmic Sobolev inequality. That is, for any smooth $g \in L^2(\mathbb{R}^n, \rho)$ (the space of square integrable functions w.r.t. the measure ρ)

$$\int g^2 \log g^2 d\rho - \int g^2 d\rho \log \left(\int g^2 d\rho \right) \leq \frac{2}{\lambda} \int |\nabla g|^2 d\rho \quad (2.12)$$

holds. If g^2 is normalized, Eq. (2.12) can be rewritten as

$$D(g^2 || \rho) \leq \frac{1}{2\lambda} I(g^2 || \rho) \quad (2.13)$$

with the Kullback-Leibler divergence D Eq. (2.2) and the relative Fisher-Information

$$I(g^2 || \rho) = \int \nabla \left| \log \frac{g^2}{\rho} \right|^2 g^2.$$

The logarithmic Sobolev inequality generalizes the Stam-Gross inequality [25, 13] where ρ is assumed to be normally distributed. The logarithmic Sobolev inequality Eq. (2.12) holds true for any distribution ρ , regardless whether it is a stationary solution of Eq. (2.9), as long as the Hessian of $-\log \rho$ fulfils the bound Eq. (2.11).

If ρ_t is a solution of the Fokker-Planck equation Eq. (2.9) with initial distribution ρ_0 s.t. $D(\rho_0 || \rho) < \infty$, we have

$$\frac{d}{dt} D(\rho_t || \rho) = - \int I(\rho_t || \rho)$$

which yields in combination with Eq. (2.13) and Gronwall's inequality

$$D(\rho_t || \rho) \leq e^{-2\lambda t} D(\rho_0 || \rho) \quad (2.14)$$

which proves an exponentially fast trend of any solution of Eq. (2.9) towards equilibrium ρ . See [19] and the references therein for more details and proofs.

3 Inferring Volatility

We define and compute the information content of returns about the volatility for stocks subject to the dynamics Eq. (2.6). In the sequel we assume

Assumption 3.1. *The Fokker-Planck equation Eq. (2.7) has a unique stationary solution ρ .*

3.1 Information calculations in stochastic volatility models

The Euler approximation of the SDE Eq. (2.6) reads

$$\begin{aligned} S_{t+\tau} - S_t &= \mu(\mathbf{v}_t)S_t\tau + f(\mathbf{v}_t)S_t\sqrt{\tau}\epsilon^0 \\ \mathbf{v}_{t+\tau} - \mathbf{v}_t &= \beta(\mathbf{v}_t)\tau + \gamma(\mathbf{v}_t)\sqrt{\tau}\epsilon \end{aligned}$$

where ϵ^0 and ϵ are 1 and n dimensional standard normal distributions s.t. $\mathbb{E}[\epsilon^0\epsilon] = (\rho_1, \dots, \rho_n)$ with the correlation coefficients ρ_i of Eq. (2.6). Defining $r_{t+\tau} = \frac{S_{t+\tau} - S_t}{S_t}$ the result is an independent return process where each $r_{t+\tau}$ is normally distributed with mean $\mu(\mathbf{v}_t)\tau$ and standard deviation $f(\mathbf{v}_t)\sqrt{\tau}$, that is,

$$\begin{aligned} r_{t+\tau} &= \mu(\mathbf{v}_t)\tau + f(\mathbf{v}_t)\sqrt{\tau}\epsilon^0 \\ \mathbf{v}_{t+\tau} - \mathbf{v}_t &= \beta(\mathbf{v}_t)\tau + \gamma(\mathbf{v}_t)\sqrt{\tau}\epsilon. \end{aligned} \quad (3.1)$$

Thus, $(r_{t+\tau}, \mathbf{v}_{t+\tau})$ are jointly generated from \mathbf{v}_t and this process is Markovian. It follows, for instance, that the past trajectory

$$\{r, \mathbf{v}\}_{t-m\tau} = (r_{t-m\tau}, \mathbf{v}_{t-m\tau}), \dots, (r_{t-\tau}, \mathbf{v}_{t-\tau}),$$

and the future one

$$\{r, \mathbf{v}\}_{t+n\tau} = (r_{t+n\tau}, \mathbf{v}_{t+n\tau}), \dots, (r_{t+\tau}, \mathbf{v}_{t+\tau})$$

provided \mathbf{v}_t form a Markov Chain 2.1

$$\{r, \mathbf{v}\}_{t-m\tau} \rightarrow \mathbf{v}_t \rightarrow \{r, \mathbf{v}\}_{t+n\tau}.$$

Note, this remains true if the process \mathbf{v}_t is not approximated. Accordingly, the Euler approximated return process is a hidden Markov process where the observation $r_{t+\tau}$ is drawn depending on \mathbf{v}_t and due to the leverage effect $\mathbf{v}_{t+\tau}$. Fig. (1) illustrates the result of the approximation. Using these conditional independences, we can approximate the information

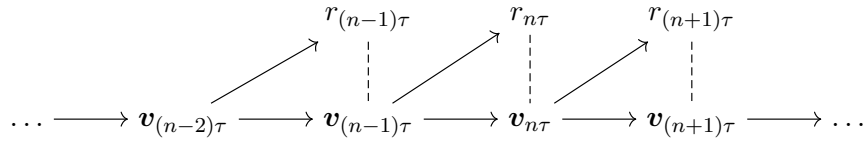


Figure 1: Euler approximation of the SDE Eq. (2.6).

$$I(f(\mathbf{v}_t)\sqrt{\tau} : \mathbf{r}_{t-n\tau}^t)$$

that historical returns

$$\mathbf{r}_{t-n\tau}^t = r_{t-n\tau}, \dots, r_t \quad (3.2)$$

provide about the volatility $f(\mathbf{v}_t)\sqrt{\tau}$

Proposition 3.2.

$$I(f(\mathbf{v}_t)\sqrt{\tau} : \mathbf{r}_{t-n\tau}^t) \leq I(\mathbf{v}_t : \mathbf{v}_{t-\tau}) + I(\mathbf{v}_t : r_t | \mathbf{v}_{t-\tau})$$

Proof. From scaling invariance of the mutual information Eq. (2.5) follows

$$I(f(\mathbf{v}_t)\sqrt{\tau} : \mathbf{r}_{t-n\tau}^t) = I(f(\mathbf{v}_t) : \mathbf{r}_{t-n\tau}^t).$$

Observe that

$$r_{t-n\tau}^t \rightarrow \mathbf{v}_t \rightarrow f(\mathbf{v}_t)$$

is a Markov Chain and the Data processing inequality 2.1 yields

$$I(f(\mathbf{v}_t) : \mathbf{r}_{t-n\tau}^t) \leq I(\mathbf{v}_t : \mathbf{r}_{t-n\tau}^t).$$

Furthermore,

$$\begin{aligned} I(\mathbf{v}_t : \mathbf{r}_{t-n\tau}^t) &\leq I(\mathbf{v}_t : \mathbf{v}_{t-\tau} \mathbf{r}_{t-n\tau}^t) \\ &= I(\mathbf{v}_t : \mathbf{v}_{t-\tau}) + I(\mathbf{v}_t : \mathbf{r}_{t-n\tau}^t | \mathbf{v}_{t-\tau}) \\ &= I(\mathbf{v}_t : \mathbf{v}_{t-\tau}) + \underbrace{I(\mathbf{v}_t : \mathbf{r}_{t-n\tau}^{t-\tau} | \mathbf{v}_{t-\tau})}_{=0} + I(\mathbf{v}_t : r_t | \mathbf{v}_{t-\tau}, \mathbf{r}_{t-n\tau}^{t-\tau}) \\ &= I(\mathbf{v}_t : \mathbf{v}_{t-\tau}) + I(\mathbf{v}_t : r_t | \mathbf{v}_{t-\tau}) \end{aligned}$$

by conditional independence $\mathbf{v}_t \perp\!\!\!\perp \mathbf{r}_{t-n\tau}^{t-\tau} | \mathbf{v}_{t-\tau}$ and $(\mathbf{v}_t, r_t) \perp\!\!\!\perp \mathbf{r}_{t-n\tau}^{t-\tau} | \mathbf{v}_{t-\tau}$. \square

An analogous argument and computation yields also an upper bound for the information past returns provide about future ones.

Proposition 3.3.

$$I(r_{t+\tau} : \mathbf{r}_{t-n\tau}^t) \leq I(r_{t+\tau} : \mathbf{v}_t)$$

Proof.

$$\begin{aligned} I(r_{t+\tau} : \mathbf{r}_{t-n\tau}^t) &\leq I(r_{t+\tau} : \mathbf{r}_{t-n\tau}^t, \mathbf{v}_t) \\ &= I(r_{t+\tau} : \mathbf{v}_t) + I(r_{t+\tau} : \mathbf{r}_{t-n\tau}^t | \mathbf{v}_t) \\ &= I(r_{t+\tau} : \mathbf{v}_t). \end{aligned}$$

\square

Hence, for computing upper bounds on the mutual informations $I(\mathbf{v}_t : \mathbf{r}_{t-n\tau}^t)$ and $I(r_{t+\tau} : \mathbf{r}_{t-n\tau}^t)$ it suffices to compute

$$\begin{aligned} I(\mathbf{v}_\tau : \mathbf{v}_0) \\ I(\mathbf{v}_\tau : r_\tau | \mathbf{v}_0) \\ I(r_\tau : \mathbf{v}_0) \end{aligned} \tag{3.3}$$

where we have set w.l.o.g. $t = \tau$ in the first two cases and $t = 0$ in the last case. Since for continuous random variables infinitely much information can be obtained (indeed, the first two terms in Eq. (3.3) diverge to ∞ if $\tau \rightarrow 0$), evaluation whether much information is gained or not from return observations needs benchmarks. Information gain may be interpreted as uncertainty reduction. Our initial uncertainty about a future return r_τ or volatility $f(\mathbf{v}_\tau)\sqrt{\tau}$ is quantified by their entropies. Prior to any observation, we think of r_τ being roughly normally distributed with standard deviation

$$\sqrt{\tau} \int f d\rho \equiv \sqrt{\tau} \sigma_f$$

i.e., the expectation of $\mathbf{v} \mapsto f(\mathbf{v})$ w.r.t the stationary distribution ρ times $\sqrt{\tau}$. Therefore, according to example 2.1, the entropy of the random variable r_τ reads

$$h(r_\tau) = \frac{1}{2} \log(2\pi e \sigma_f^2 \tau).$$

Since \mathbf{v}_τ being drawn from the stationary distribution ρ with entropy $h(\mathbf{v}_\tau)$ the volatility $f(\mathbf{v}_\tau)\sqrt{\tau}$ is stationary distributed with entropy $h(f(\mathbf{v}_\tau)\sqrt{\tau})$. If f is a diffeomorphic coordinate change of a one-dimensional process v_τ the entropy of the volatility is

$$h(f(v_\tau)\sqrt{\tau}) = h(v_\tau) + \int \log |f'(v)\sqrt{\tau}| d\rho(v).$$

Similar to the example 2.2 we can compute how much information would be required in order to know r_τ and $f(\mathbf{v}_\tau)\sqrt{\tau}$ up to some precision. We think of r_τ being quoted in percent with precision 0.01 on a daily basis whereas $f(\mathbf{v}_\tau)\sqrt{\tau}$ is quoted like the VIX, the volatility index of the S&P 500. The VIX is quoted in percentage points as well up to precision 0.01 and translates, roughly, to the expected movement (with the assumption of a 68% likelihood, i.e., one standard deviation) in the S&P 500 index over the next 30-day period, which is then annualized. This fits into our setting as follows. Assume we have an annual time scale for the SDE Eq. (2.6) and read off daily returns (a situation we encounter in subsection 4.1). Then $\tau = 1/252$ (there are 252 trading days a year) and the volatility of the daily returns in the Euler approximation is $f(\mathbf{v}_\tau)\sqrt{\tau}$. However, since volatility is quoted on an annual basis, we are rather interested in the random variable $f(\mathbf{v}_\tau)$ than $f(\mathbf{v}_\tau)\sqrt{\tau}$. According to example 2.2, we have to compute the differences

$$\begin{aligned} G_r &\equiv h(r_\tau) - \frac{1}{2} \log(2\pi e \sigma_M^2) = \log\left(\frac{\sigma_f \sqrt{\tau}}{\sigma_M}\right) \\ G_{f(\mathbf{v})} &\equiv h(f(\mathbf{v}_\tau)) - \frac{1}{2} \log(2\pi e \sigma_M^2) \end{aligned} \quad (3.4)$$

where $\sigma_M = \frac{1}{4}10^{-4}$ because we want precision up to 0.01 percent. If we are in the one-dimensional setting with a diffeomorphic function f then the second identity reads

$$G_{f(v)} = h(v_\tau) + \int \log |f'| d\rho - \frac{1}{2} \log(2\pi e \sigma_M^2).$$

However, if the parameters of the SDE Eq. (2.6) are quoted according to a daily time scale, we have $\tau = 1$ and we have to rescale the volatility $f(\mathbf{v}_\tau)$, which is now the one on a daily basis, to $f(\mathbf{v}_\tau)\sqrt{252}$ in order to make the result comparable with the VIX. Hence, we obtain the equations

$$\begin{aligned} G_r &\equiv h(r_\tau) - \frac{1}{2} \log(2\pi e \sigma_M^2) = \log\left(\frac{\sigma_f}{\sigma_M}\right) \\ G_{f(\mathbf{v})} &\equiv h(f(\mathbf{v}_\tau)) + \frac{1}{2} \log(252) - \frac{1}{2} \log(2\pi e \sigma_M^2) \end{aligned} \quad (3.5)$$

with $\sigma_M = \frac{1}{4}10^{-4}$.

3.2 General Solution

We derive upper bounds and proxies for the mutual informations Eq. (3.3) in the subsequent theorems.

Definition 3.1. We define the mutual information proxy

$$\mathcal{I}(\mathbf{v}_\tau, \mathbf{v}_0) = h(\mathbf{v}_0) - \frac{1}{2} \int \log (2^n \pi^n e \det (\gamma(\mathbf{v}_0) \gamma^T(\mathbf{v}_0)) \tau^n) d\rho(\mathbf{v}_0)$$

where

$$h(\mathbf{v}_0) = - \int \log \rho(\mathbf{v}_0) d\rho(\mathbf{v}_0)$$

denotes the entropy of the stationary distribution.

Theorem 3.4. In the Euler approximation scheme Fig. (1) we obtain

$$I(\mathbf{v}_\tau : \mathbf{v}_0) = \mathcal{I}(\mathbf{v}_\tau, \mathbf{v}_0) + \mathcal{O}(\tau^2)$$

and

$$I(r_\tau : \mathbf{v}_\tau | \mathbf{v}_0) = -\frac{1}{2} \log \left(\frac{1}{n^2} \sum_{i=1}^n 1 - \rho_i^2 \right)$$

for $i = 1, \dots, n$ where $\rho_i = dW_t^i dW_t^0$ are the correlations in the SDE Eq. (2.6).

Proof. We introduce the Fokker-Planck operator

$$L_{FP}(\mathbf{v}) = -\nabla \beta(\mathbf{v}) + \frac{1}{2} \sum_{i,j=1}^n \partial_i \partial_j \gamma^2(\mathbf{v}).$$

Then, the normal distribution obtained from the Euler approximation can be also rewritten as

$$p(\mathbf{v}_\tau | \mathbf{v}_0) = e^{\tau L_{FP}(\mathbf{v}_0)} \delta_{\mathbf{v}_\tau - \mathbf{v}_0} = (1 + \tau L_{FP}(\mathbf{v}_0) + \mathcal{O}(\tau^2)) \delta_{\mathbf{v}_\tau - \mathbf{v}_0}$$

where δ_x denotes the delta distribution at x . This easily follows from considering the characteristic functions on both sides, but see also chapter 4 in [22] for a proof. The Kramer-Moyal Forward Expansion for the actual transition probability $q(\mathbf{v}_\tau | \mathbf{v}_0)$, i.e., the solution of the Fokker-Planck equation Eq. (2.7) with initial condition \mathbf{v}_0 for $\tau = 0$, is as well

$$q(\mathbf{v}_\tau | \mathbf{v}_0) = (1 + \tau L_{FP}(\mathbf{v}_0) + \mathcal{O}(\tau^2)) \delta_{\mathbf{v}_\tau - \mathbf{v}_0}.$$

Hence

$$\begin{aligned} p(\mathbf{v}_\tau | \mathbf{v}_0) \rho(\mathbf{v}_0) &= (p(\mathbf{v}_\tau | \mathbf{v}_0) - q(\mathbf{v}_\tau | \mathbf{v}_0)) \rho(\mathbf{v}_0) + q(\mathbf{v}_\tau | \mathbf{v}_0) \rho(\mathbf{v}_0) \\ &= \mathcal{O}(\tau^2) \delta_{\mathbf{v}_\tau - \mathbf{v}_0} \rho(\mathbf{v}_0) + q(\mathbf{v}_\tau | \mathbf{v}_0) \rho(\mathbf{v}_0) \\ &= \mathcal{O}(\tau^2) \rho(\mathbf{v}_\tau) + q(\mathbf{v}_\tau | \mathbf{v}_0) \rho(\mathbf{v}_0). \end{aligned}$$

Since the Fokker-Planck equation Eq. (2.7) is linear, any superposition of solutions is a solution as well. Thus,

$$q(\mathbf{v}_\tau) = \int q(\mathbf{v}_\tau | \mathbf{v}_0) d\rho(\mathbf{v}_0)$$

is the solution with initial condition ρ , i.e., the stationary solution of Eq. (2.7) and therefore stationary as well and we obtain

$$p(\mathbf{v}_\tau) = \int p(\mathbf{v}_\tau | \mathbf{v}_0) d\rho(\mathbf{v}_0) = \int (\mathcal{O}(\tau^2) \rho(\mathbf{v}_\tau) + q(\mathbf{v}_\tau | \mathbf{v}_0)) d\rho(\mathbf{v}_0) = (1 + \mathcal{O}(\tau^2)) \rho(\mathbf{v}_\tau).$$

Therefore, the transformation rule Eq. (2.1) for the entropy yields

$$-\int \log p(\mathbf{v}_\tau) d\rho(\mathbf{v}_\tau) = -\int \log \rho(\mathbf{v}_\tau) d\rho(\mathbf{v}_\tau) + \int \log(1 + \mathcal{O}(\tau^2)) d\rho(\mathbf{v}_\tau) = h(\mathbf{v}_0) + \mathcal{O}(\tau^2)$$

Hence, from example 2.1 and the fact that $p(\mathbf{v}_\tau|\mathbf{v}_0)$ is normally distributed with covariance matrix $\gamma(\mathbf{v})\gamma^T(\mathbf{v})\tau$ we obtain

$$\begin{aligned} I(\mathbf{v}_\tau : \mathbf{v}_0) &= h(\mathbf{v}_\tau) - h(\mathbf{v}_\tau|\mathbf{v}_0) \\ &= h(\mathbf{v}_0) + \mathcal{O}(\tau^2) + \int p(\mathbf{v}_\tau|\mathbf{v}_0) \log p(\mathbf{v}_\tau|\mathbf{v}_0) d\mathbf{v}_\tau d\rho(\mathbf{v}_0) \\ &= h(\mathbf{v}_0) + \mathcal{O}(\tau^2) - \frac{1}{2} \int \log(2^n \pi^n e \det(\gamma(\mathbf{v}_0)\gamma^T(\mathbf{v}_0)) \tau^n) d\rho(\mathbf{v}_0) \\ &= \mathcal{I}(\mathbf{v}_\tau, \mathbf{v}_0) + \mathcal{O}(\tau^2) \end{aligned}$$

The second mutual information reads

$$I(r_\tau : \mathbf{v}_\tau|\mathbf{v}_0) = h(r_\tau|\mathbf{v}_0) - h(r_\tau|\mathbf{v}_\tau, \mathbf{v}_0).$$

r_τ is normally distributed with variance $f(\mathbf{v}_0)^2\tau$ in the Euler approximation scheme of Eq. (3.1). Hence

$$h(r_\tau|\mathbf{v}_0) = \frac{1}{2} \int \log(2\pi e f(\mathbf{v}_0)^2\tau) d\rho(\mathbf{v}_0).$$

Since a stationary solution exists, the boundary 0 is reflecting and therefore we have almost surely $|\mathbf{v}_0|^2 > 0$ and due to uniform ellipticity of Eq. (2.7) the matrix $\gamma(\mathbf{v}_0)$ is invertible and we obtain

$$\boldsymbol{\varepsilon} = \frac{1}{\sqrt{\tau}} \gamma(\mathbf{v}_0)^{-1} (\mathbf{v}_\tau - \beta(\mathbf{v}_0)\tau)$$

Since $\mathbb{E}[\boldsymbol{\varepsilon}^0 \boldsymbol{\varepsilon}] = (\rho_1, \dots, \rho_n) = \boldsymbol{\rho}$ in Eq. (3.1), the return reads

$$\begin{aligned} r_\tau &= \mu(\mathbf{v}_0)\tau + f(\mathbf{v}_0)\sqrt{\tau}\boldsymbol{\varepsilon}^0 \\ &= \mu(\mathbf{v}_0)\tau + \frac{f(\mathbf{v}_0)\sqrt{\tau}}{n} \sum_{i=1}^n \varepsilon^0_i \\ &= \mu(\mathbf{v}_0)\tau + \frac{f(\mathbf{v}_0)\sqrt{\tau}}{n} \sum_{i=1}^n \rho_i \varepsilon^0_i + \sqrt{1 - \rho_i^2} z^i \\ &= \mu(\mathbf{v}_0)\tau + \frac{f(\mathbf{v}_0)\sqrt{\tau}}{n} \sum_{i=1}^n \sqrt{1 - \rho_i^2} z^i + \frac{f(\mathbf{v}_0)\sqrt{\tau}}{n} \langle \boldsymbol{\rho}, \boldsymbol{\varepsilon} \rangle \\ &= \mu(\mathbf{v}_0)\tau + \frac{f(\mathbf{v}_0)\sqrt{\tau}}{n} \sum_{i=1}^n \sqrt{1 - \rho_i^2} z^i + \frac{f(\mathbf{v}_0)}{n} \langle \boldsymbol{\rho}, \gamma(\mathbf{v}_0)^{-1} (\mathbf{v}_\tau - \beta(\mathbf{v}_0)\tau) \rangle \end{aligned}$$

where z^i are mutually independent, standard normal random variables which are also independent from $\boldsymbol{\varepsilon}$. $\langle \cdot, \cdot \rangle$ denotes the standard scalar product. Hence, r_τ provided \mathbf{v}_τ and \mathbf{v}_0 is normally distributed with mean

$$\mu(\mathbf{v}_0)\tau + \frac{f(\mathbf{v}_0)}{n} \langle \boldsymbol{\rho}, \gamma(\mathbf{v}_0)^{-1} (\mathbf{v}_\tau - \beta(\mathbf{v}_0)\tau) \rangle$$

and variance

$$\frac{f(\mathbf{v}_0)^2\tau}{n^2} \sum_{i=1}^n 1 - \rho_i^2$$

and therefore

$$h(r_\tau | \mathbf{v}_\tau, \mathbf{v}_0) = \frac{1}{2} \int \log \left(2\pi e \frac{f(\mathbf{v}_0)^{2\tau}}{n^2} \sum_{i=1}^n 1 - \rho_i^2 \right) d\rho(\mathbf{v}_0).$$

Hence,

$$\begin{aligned} I(r_\tau : \mathbf{v}_\tau | \mathbf{v}_0) &= \frac{1}{2} \int \log (2\pi e f(\mathbf{v}_0)^{2\tau}) d\rho(\mathbf{v}_0) - \frac{1}{2} \int \log \left(2\pi e \frac{f(\mathbf{v}_0)^{2\tau}}{n^2} \sum_{i=1}^n 1 - \rho_i^2 \right) d\rho(\mathbf{v}_0) \\ &= -\frac{1}{2} \int \log \left(\frac{1}{n^2} \sum_{i=1}^n 1 - \rho_i^2 \right) d\rho(\mathbf{v}_0) = -\frac{1}{2} \log \left(\frac{1}{n^2} \sum_{i=1}^n 1 - \rho_i^2 \right) \end{aligned}$$

□

Remark 3.1. 1. The proxy $\mathcal{I}(\mathbf{v}_\tau, \mathbf{v}_0)$ on the mutual information $I(\mathbf{v}_\tau : \mathbf{v}_0)$ does depend on the parametrization of the vector \mathbf{v}_τ .

2. If the dimension n of the volatility process is one-dimensional, and f is an diffeomorphism, we can substitute \mathbf{v}_τ and \mathbf{v}_0 by $f(\mathbf{v}_\tau)$ and $f(\mathbf{v}_0)$ in theorem 3.4, respectively, due to the invariance of mutual and conditional mutual information.

In the remaining of this subsection we assume that the stationary distribution ρ fulfils the logarithmic Sobolev inequality Eq. (2.12) with parameter $\lambda > 0$. Furthermore, we introduce the *Fisher information matrix*

$$\mathcal{I}(r_\tau; \mathbf{v}_0)_{ij} = \int \partial_{v_0^i} \log p(r_\tau | \mathbf{v}_0) \partial_{v_0^j} \log p(r_\tau | \mathbf{v}_0) p(r_\tau | \mathbf{v}_0) dr_\tau$$

with $i, j = 1, \dots, n$ for the conditional density $p(r_\tau | \mathbf{v}_0)$ and the Fisher information

$$\mathcal{I}(r_\tau; f(\mathbf{v}_0)) = \int \left(\frac{d}{dy} \log p(r_\tau | y) \Big|_{y=f(\mathbf{v}_0)} \right)^2 p(r_\tau | f(\mathbf{v}_0)) dr_\tau$$

for the conditional density $p(r_\tau | f(\mathbf{v}_0))$. The Fisher information can be interpreted as the information about \mathbf{v}_0 and $f(\mathbf{v}_0)$, respectively, that is present in the return r_τ . See [9] for a detailed discussion of the properties and the interpretation of the Fisher information.

Theorem 3.5. Let ρ fullfill Eq. (2.12) with $\lambda > 0$. Then

$$I(r_\tau : \mathbf{v}_0) \leq \frac{1}{2\lambda} \int \text{Tr}(\mathcal{I}(r_\tau; \mathbf{v}_0)) d\rho(\mathbf{v}_0). \quad (3.6)$$

If there is a factor, that is, a smooth function $\tilde{\mu} : \mathbb{R} \rightarrow \mathbb{R}$ s.t.

$$\tilde{\mu} \circ f(\mathbf{v}_0) = \mu(\mathbf{v}_0),$$

then

$$\begin{aligned} \text{Tr}(\mathcal{I}(r_\tau; \mathbf{v}_0)) &= \mathcal{I}(r_\tau; f(\mathbf{v}_0)) |\nabla f(\mathbf{v}_0)|^2 \\ &= ((\tilde{\mu}' \circ f(\mathbf{v}_0))^2 \tau + 2) |\nabla \log f(\mathbf{v}_0)|^2 \end{aligned} \quad (3.7)$$

Proof. If we set $g = \sqrt{p(r_\tau|\mathbf{v}_0)}$ in Eq. (2.12), we get

$$\begin{aligned}
I(r_\tau : \mathbf{v}_0) &= H(r_\tau) - H(r_\tau|\mathbf{v}_0) \\
&= - \int p(r_\tau) \log p(r_\tau) dr_\tau + \int p(r_\tau, \mathbf{v}_0) \log p(r_\tau|\mathbf{v}_0) d\mathbf{v}_0 dr_\tau \\
&= - \int \int p(r_\tau|\mathbf{v}_0) d\rho(\mathbf{v}_0) \log \left(\int p(r_\tau|\mathbf{v}_0) d\rho(\mathbf{v}_0) \right) + \int p(r_\tau|\mathbf{v}_0) \log p(r_\tau|\mathbf{v}_0) d\rho(\mathbf{v}_0) dr_\tau \\
&\leq \frac{2}{\lambda} \int |\nabla \sqrt{p(r_\tau|\mathbf{v}_0)}|^2 d\rho(\mathbf{v}_0) dr_\tau \\
&= \frac{1}{2\lambda} \int \sum_{i=1}^n \frac{\partial_{v_0^i} p(r_\tau|\mathbf{v}_0)^2}{p(r_\tau|\mathbf{v}_0)} d\rho(\mathbf{v}_0) dr_\tau \\
&= \frac{1}{2\lambda} \int \sum_{i=1}^n \left(\partial_{v_0^i} \log p(r_\tau|\mathbf{v}_0) \right)^2 p(r_\tau|\mathbf{v}_0) d\rho(\mathbf{v}_0) dr_\tau \\
&= \frac{1}{2\lambda} \int \sum_{i=1}^n \mathcal{I}(r_\tau; \mathbf{v}_0)_{ii} d\rho(\mathbf{v}_0) \\
&= \frac{1}{2\lambda} \int \text{Tr}(\mathcal{I}(r_\tau; \mathbf{v}_0)) d\rho(\mathbf{v}_0)
\end{aligned}$$

If $\mu(\mathbf{v}_0) = \tilde{\mu} \circ f(\mathbf{v}_0)$ the second identity follows from

$$p(r_\tau|\mathbf{v}_0) = p(r_\tau|f(\mathbf{v}_0)),$$

that is, the conditional distribution only depends on $f(\mathbf{v}_0)$, and the chain rule

$$\partial_{v_0^i} \log p(r_\tau|\mathbf{v}_0) = \partial_{v_0^i} f(\mathbf{v}_0) \left. \frac{d}{dy} \log p(r_\tau|y) \right|_{y=f(\mathbf{v}_0)}$$

for $i = 1, \dots, n$. Finally, we compute

$$\begin{aligned}
\frac{d}{dy} \log p(r_\tau|y) &= \frac{d}{dy} \log \left(\frac{1}{\sqrt{2\pi\tau y}} e^{-\frac{(r_\tau - \tilde{\mu}(y)\tau)^2}{2y^2\tau}} \right) \\
&= \frac{d}{dy} \left(-\log y - \frac{(r_\tau - \tilde{\mu}(y)\tau)^2}{2y^2\tau} \right) \\
&= -\frac{1}{y} - \frac{1}{2\tau} \frac{-2(r_\tau - \tilde{\mu}(y)\tau)\tilde{\mu}'(y)\tau y^2 - 2y(r_\tau - \tilde{\mu}(y)\tau)^2}{y^4} \\
&= -\frac{1}{y} - \frac{1}{\tau} \frac{(r_\tau - \tilde{\mu}(y)\tau)\tilde{\mu}'(y)\tau y + (r_\tau - \tilde{\mu}(y)\tau)^2}{y^3}
\end{aligned}$$

which yields

$$\begin{aligned}
& \int \left(\frac{1}{\tau} \frac{(r_\tau - \tilde{\mu}(y)\tau)\tilde{\mu}'(y)\tau y + (r_\tau - \tilde{\mu}(y)\tau)^2}{y^3} - \frac{1}{y} \right)^2 \frac{1}{\sqrt{2\pi\tau y}} e^{-\frac{(x - \tilde{\mu}(y)\tau)^2}{2y^2\tau}} dr_\tau \\
&= \int \left(\left(\frac{1}{\tau} \frac{(r_\tau - \tilde{\mu}(y)\tau)\tilde{\mu}'(y)\tau y + (r_\tau - \tilde{\mu}(y)\tau)^2}{y^3} \right)^2 + \right. \\
&\quad \left. - \frac{2}{y} \frac{1}{\tau} \frac{(r_\tau - \tilde{\mu}(y)\tau)\tilde{\mu}'(y)\tau y + (r_\tau - \tilde{\mu}(y)\tau)^2}{y^3} + \frac{1}{y^2} \right) \frac{1}{\sqrt{2\pi\tau y}} e^{-\frac{(x - \tilde{\mu}(y)\tau)^2}{2y^2\tau}} dr_\tau \\
&= \int \left(\frac{1}{\tau^2} \frac{((r_\tau - \tilde{\mu}(y)\tau)\tilde{\mu}'(y)\tau)^2}{y^4} + \frac{1}{\tau^2} \frac{2(r_\tau - \tilde{\mu}(y)\tau)^3 \tilde{\mu}'(y)\tau}{y^5} + \frac{1}{\tau^2} \frac{(r_\tau - \tilde{\mu}(y)\tau)^4}{y^6} + \right. \\
&\quad \left. - \frac{2}{y} \frac{1}{\tau} \frac{(r_\tau - \tilde{\mu}(y)\tau)\tilde{\mu}'(y)\tau y + (r_\tau - \tilde{\mu}(y)\tau)^2}{y^3} + \frac{1}{y^2} \right) \frac{1}{\sqrt{2\pi\tau y}} e^{-\frac{(x - \tilde{\mu}(y)\tau)^2}{2y^2\tau}} dr_\tau \\
&= \frac{\tilde{\mu}'(y)^2 \tau}{y^2} + \frac{3}{y^2} - \frac{2}{y^2} + \frac{1}{y^2} \\
&= \frac{\tilde{\mu}'(y)^2 \tau}{y^2} + \frac{2}{y^2}
\end{aligned}$$

Hence, if we substitute y by $f(\mathbf{v}_0)$, we obtain

$$\mathcal{I}(r_\tau; f(\mathbf{v}_0)) = \frac{(\tilde{\mu}' \circ f(\mathbf{v}_0))^2 \tau}{f(\mathbf{v}_0)^2} + \frac{2}{f(\mathbf{v}_0)^2}$$

and therefore

$$\begin{aligned}
\mathcal{I}(r_\tau; f(\mathbf{v}_0)) |\nabla f(\mathbf{v}_0)|^2 &= \left(\frac{(\tilde{\mu}' \circ f(\mathbf{v}_0))^2 \tau}{f(\mathbf{v}_0)^2} + \frac{2}{f(\mathbf{v}_0)^2} \right) |\nabla f(\mathbf{v}_0)|^2 \\
&= ((\tilde{\mu}' \circ f(\mathbf{v}_0))^2 \tau + 2) |\nabla \log f(\mathbf{v}_0)|^2
\end{aligned}$$

□

Remark 3.2. 1. If there is a constant factor $\tilde{\mu}$, the upper bound does no longer depend on τ . Hence, previous returns $\mathbf{r}_{t-n\tau}^t$ do not provide more information about the future return $r_{t+\tau}$ than suggested by the τ -independent upper bound derived in theorem 3.5 even if $\tau \rightarrow 0$ and $n \rightarrow \infty$.

2. The inequality Eq. (3.6) and identity Eq. (3.7) hold true in general, i.e., for any smooth conditional density $p(r_\tau | \mathbf{v}_0)$.
3. Inequality Eq. (3.6) has a nice interpretation: the mutual information $I(r_\tau : \mathbf{v}_0)$ is bounded by the weighted Fisher information $\mathcal{I}(r_\tau; \mathbf{v}_0)$ every return yields on the particular volatility \mathbf{v}_0 with weights $\rho(\mathbf{v}_0)$.
4. The upper bound Eq. (3.6) for the mutual information depends on the parametrization of \mathbf{v}_0 whereas the mutual information $I(r_\tau : \mathbf{v}_0)$ itself is invariant.
5. Recall, the proof of theorem 3.4 yields

$$p(\mathbf{v}_\tau) = \int p(\mathbf{v}_\tau | \mathbf{v}_0) d\rho(\mathbf{v}_0) = \rho(\mathbf{v}_\tau)$$

where $p(\mathbf{v}_\tau|\mathbf{v}_0)$ is the solution of the Fokker-Planck equation Eq. (2.7) with initial condition $\delta_{\mathbf{v}_0}$. According to Eq. (2.3) the mutual information $I(\mathbf{v}_\tau : \mathbf{v}_0)$ reads

$$\begin{aligned} I(\mathbf{v}_\tau : \mathbf{v}_0) &= D(p(\mathbf{v}_\tau, \mathbf{v}_0) || p(\mathbf{v}_\tau)p(\mathbf{v}_0)) \\ &= \int p(\mathbf{v}_\tau, \mathbf{v}_0) \log \frac{p(\mathbf{v}_\tau, \mathbf{v}_0)}{p(\mathbf{v}_0)p(\mathbf{v}_\tau)} d\mathbf{v}_\tau d\mathbf{v}_0 \\ &= \int p(\mathbf{v}_\tau|\mathbf{v}_0) \log \frac{p(\mathbf{v}_\tau|\mathbf{v}_0)}{\rho(\mathbf{v}_\tau)} d\mathbf{v}_\tau d\rho(\mathbf{v}_0) \\ &= \int D(\rho_\tau || \rho) d\rho \end{aligned}$$

where $\rho_\tau = p(\mathbf{v}_\tau|\mathbf{v}_0)$ is the solution of the Fokker-Planck equation Eq. (2.7) with initial condition $\rho_0 = \delta_{\mathbf{v}_0}$. Formally $D(\delta_{\mathbf{v}_0} || \rho) = \infty$. Informally, one could argue, volatility is not known precisely and we approximate the Dirac distribution $\delta_{\mathbf{v}_0}$ by an initial distribution $\rho_{\mathbf{v}_0}$ with mean \mathbf{v}_0 and small variance s.t. $D(\rho_{\mathbf{v}_0} || \rho) < \infty$. In the case that the stationary distribution fulfils the logarithmic Sobolev inequality Eq. (2.12) we obtain from Eq. (2.14)

$$I(\mathbf{v}_\tau : \mathbf{v}_0) \leq e^{-2\lambda\tau} \int D(\rho_{\mathbf{v}_0} || \rho) d\rho$$

Hence, assuming that the integral on the right of the inequality is finite, we have an informal argument that the mutual information $I(\mathbf{v}_\tau : \mathbf{v}_0)$ decreases exponentially in time.

6. As in remark 3.1, if the volatility vector v_0 is one dimensional and f diffeomorphic, $I(r_\tau, v_0)$ can be replaced by $I(r_\tau, f(v_0))$ in theorem 3.5.

4 Stochastic Volatility Models

We compute the mutual informations Eq. (3.3) for various stochastic volatility models.

4.1 Mean Reverting One Factor Models

We follow [8] and consider six stochastic volatility models of the form

$$\begin{aligned} dS_t &= rS_t dt + \sqrt{v_t} S_t dW_t^0 \\ dv_t &= \gamma v_t^a (\theta - v_t) dt + \kappa v_t^b dW_t^1 \end{aligned}$$

with $dW_t^1 dW_t^0 = \rho dt$ where $a \in \{0, 1\}$ and $b \in \{1/2, 1, 3/2\}$. Jones [15] proves that for $a = 0$ and $b > 1$, there are unique stationary solutions. Analogous arguments yield stationary solutions (and therefore also uniqueness, according to lemma 2.2) in the cases $b = 1; a = 0, 1$ and $b = 1/2; a = 0$. In these cases we compute the stationary distribution and upper bounds or proxies for the mutual informations Eq. (3.3). If $b = 1/2$ and $a = 1$, we can prove that there is no stationary solution. We start with

Lemma 4.1. *If there is a stationary distribution ρ for v we have*

$$\begin{aligned} \mathcal{I}(v_\tau, v_0) &= h(v_0) - \frac{1}{2} \log(2e\pi\kappa^2\tau) - b \int \log v_0 d\rho(v_0) \\ I(r_\tau : v_\tau | v_0) &= -\frac{1}{2} \log(1 - \rho^2) \end{aligned}$$

Proof. According to theorem 3.4 we have

$$\begin{aligned}\mathcal{I}(v_\tau, v_0) &= h(v_0) - \frac{1}{2} \int \log \left(2\pi e \kappa^2 v_0^{2b} \tau \right) d\rho(v_0) \\ &= h(v_0) - \frac{1}{2} \log (2\pi e \kappa^2 \tau) - b \int \log v_0 d\rho(v_0).\end{aligned}$$

and the second equality follows directly with $n = 1$. \square

In order to compute the stationary solutions for the various models explicitly we transform the SDE for v_t into the gradient flow Eq. (2.8) form.

Lemma 4.2. *Define*

$$\begin{aligned}g(v) &= \frac{\sqrt{2}}{\kappa(1-b)} v^{1-b} \quad \text{if } b \neq 1 \\ g(v) &= \frac{\sqrt{2}}{\kappa} \log v \quad \text{if } b = 1\end{aligned}$$

and $\sigma = g(v)$. Then

$$d\sigma_t = \sqrt{2} dW_t^1 + V'(\sigma_t) dt$$

with

$$\begin{aligned}V'(\sigma) &= \sqrt{2} \left\{ -\frac{\gamma}{\kappa} \left(\theta \left(\frac{\sigma \kappa (1-b)}{\sqrt{2}} \right)^{\frac{a-b}{1-b}} - \left(\frac{\sigma \kappa (1-b)}{\sqrt{2}} \right)^{\frac{a-b+1}{1-b}} \right) + \frac{b}{(1-b)\sqrt{2}} \frac{1}{\sigma} \right\} \quad \text{if } b \neq 1 \\ V'(\sigma) &= \sqrt{2} \left\{ -\frac{\gamma}{\kappa} \left(\theta e^{\frac{(a-1)\kappa}{\sqrt{2}} \sigma} - e^{\frac{a\kappa}{\sqrt{2}} \sigma} \right) + \frac{\kappa}{2} \right\} \quad \text{if } b = 1\end{aligned}$$

Proof. For $b \neq 1$ we have

$$\begin{aligned}g'(v) &= \frac{\sqrt{2}}{\kappa} \frac{1}{v^b} \\ g''(v) &= -\frac{b\sqrt{2}}{\kappa} \frac{1}{v^{1+b}} \\ v &= \left(\frac{\sigma \kappa (1-b)}{\sqrt{2}} \right)^{\frac{1}{1-b}}\end{aligned}$$

and Itô's formula yields

$$\begin{aligned}d\sigma &= g'(v_t) dv_t + \frac{g''(v_t)}{2} dv_t dv_t \\ &= \frac{\sqrt{2}}{\kappa} \frac{1}{v_t^b} \left(\gamma v_t^a (\theta - v_t) dt + \kappa v_t^b dW_t^1 \right) - \frac{b}{\kappa \sqrt{2}} \frac{1}{v_t^{1+b}} \left(\kappa v_t^b \right)^2 dt \\ &= \sqrt{2} dW_t^1 + \frac{\sqrt{2}\gamma}{\kappa} v^{a-b} (\theta - v_t) dt - \frac{b\kappa}{\sqrt{2}} v^{b-1} dt \\ &= \sqrt{2} dW_t^1 + \sqrt{2} \left\{ \frac{\gamma}{\kappa} \left(\frac{\sigma \kappa (1-b)}{\sqrt{2}} \right)^{\frac{a-b}{1-b}} \left(\theta - \left(\frac{\sigma \kappa (1-b)}{\sqrt{2}} \right)^{\frac{1}{1-b}} \right) - \frac{b}{\sqrt{2}(1-b)} \frac{1}{\sigma} \right\} dt \\ &= \sqrt{2} dW_t^1 + \sqrt{2} \left\{ \frac{\gamma}{\kappa} \left(\theta \left(\frac{\sigma \kappa (1-b)}{\sqrt{2}} \right)^{\frac{a-b}{1-b}} - \left(\frac{\sigma \kappa (1-b)}{\sqrt{2}} \right)^{\frac{a-b+1}{1-b}} \right) - \frac{b}{\sqrt{2}(1-b)} \frac{1}{\sigma} \right\} dt.\end{aligned}$$

For $b = 1$ we have

$$\begin{aligned} g'(v) &= \frac{\sqrt{2}}{\kappa} \frac{1}{v} \\ g''(v) &= -\frac{\sqrt{2}}{\kappa} \frac{1}{v^2} \\ v &= e^{\kappa\sigma/\sqrt{2}} \end{aligned}$$

and Itô's formula yields

$$\begin{aligned} d\sigma &= g'(v_t)dv_t + \frac{g''(v_t)}{2}dv_tdv_t \\ &= \frac{\sqrt{2}}{\kappa} \frac{1}{v_t} (\gamma v_t^a(\theta - v_t)dt + \kappa v_t dW_t^1) - \frac{1}{\kappa\sqrt{2}} \frac{1}{v_t^2} (\kappa v_t)^2 dt \\ &= \sqrt{2}dW_t^1 + \frac{\sqrt{2}\gamma}{\kappa} v_t^{a-1}(\theta - v_t)dt - \frac{\kappa}{\sqrt{2}} dt \\ &= \sqrt{2}dW_t^1 + \sqrt{2} \left\{ \frac{\gamma}{\kappa} \left(\theta e^{\frac{(a-1)\kappa}{\sqrt{2}}\sigma} - e^{\frac{a\kappa}{\sqrt{2}}\sigma} \right) - \frac{\kappa}{2} \right\} dt \end{aligned}$$

□

In order to apply theorem 3.5 the following lemma turns out to be useful.

Lemma 4.3. *The function*

$$\rho(\sigma) = \frac{2\beta^\alpha}{\Gamma(\alpha)} \sigma^{2\alpha-1} e^{-\beta\sigma^2}$$

defines a probability density for all $\alpha, \beta > 0$. σ^2 is Gamma distributed with shape α and rate β . The density ρ fulfils the logarithmic Sobolev inequality Eq. (2.12) with

$$\lambda = 2\beta.$$

Furthermore, assume the process v_t has a stationary distribution $\rho(v)$ and $v = g(\sigma)$ with

$$g(\sigma) = \eta\sigma^b \quad b \in \{-2, 2\}, \eta > 0$$

and $\alpha > 1$, then

$$I(r_\tau : v_0) \leq \frac{1}{2(\alpha - 1)}.$$

Proof. We first check that $u = h(\sigma) = \sigma^2$ is Gamma distributed with shape α and rate β . The transformation rule for probability densities yields

$$\begin{aligned} \rho(u) &= \rho(h^{-1}(u)) \frac{1}{h'(h^{-1}(u))} \\ &= \rho(\sqrt{u}) \frac{1}{2\sqrt{u}} \\ &= \frac{2\beta^\alpha}{\Gamma(\alpha)} u^{\alpha-1/2} e^{-\beta u} \frac{1}{2\sqrt{u}} \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} u^{\alpha-1} e^{-\beta u} \end{aligned}$$

which is the Gamma distribution with shape α and rate β . Hence, $\rho(\sigma)$ is a density as well. Furthermore, if we write $\rho(\sigma) = e^{-V(\sigma)}$ we obtain

$$V(\sigma) = \beta\sigma^2 - (2\alpha - 1)\log \sigma + Z$$

for a constant Z . Hence

$$V''(\sigma) = 2\beta$$

is bounded from below by $2\beta > 0$ and ρ fulfils the logarithmic Sobolev inequality. Since mutual information is scaling invariant we have for $\sigma_0 = g^{-1}(v_0)$

$$I(r_\tau : v_0) = I(r_\tau : \sigma_0).$$

We apply theorem 3.5 and get

$$\begin{aligned} \lambda &= 2\beta \\ f(\sigma_0) &= \sqrt{g(g^{-1}(v_0))} = \sqrt{\eta\sigma_0^b} = \sqrt{\eta}\sigma_0^{b/2}; \quad b \in \{-2, 2\} \\ \tilde{\mu}' &= 0 \end{aligned} \tag{4.1}$$

where the last identity follows from the fact that the drift term is constant. Hence, we obtain

$$\begin{aligned} I(r_\tau : \sigma_0) &\leq \frac{1}{4\beta} \int 2 \left| \frac{d}{d\sigma} \log f(\sigma_0) \right|^2 d\rho(\sigma_0) \\ &= \frac{1}{2\beta} \int \frac{1}{\sigma_0^2} d\rho(\sigma_0) \\ &= \frac{1}{2\beta} \int \frac{1}{\sigma_0^2} \frac{2\beta^\alpha}{\Gamma(\alpha)} \sigma_0^{2\alpha-1} e^{-\beta\sigma_0^2} d\sigma_0 \\ &= \frac{1}{2} \int \frac{2\beta^{\alpha-1}}{\Gamma(\alpha)} \sigma_0^{2(\alpha-1)} e^{-\beta\sigma_0^2} d\sigma_0 \\ &= \frac{\Gamma(\alpha-1)}{2\Gamma(\alpha)} \\ &= \frac{1}{2(\alpha-1)} \end{aligned}$$

where the second identity follows from Eq. (4.1). □

Remark 4.1. In the cases $b = 1/2$ and $b = 3/2$ we have from lemma 4.2

$$v = \left(\frac{\sigma\kappa(1-b)}{\sqrt{2}} \right)^{\frac{1}{1-b}},$$

that is, we have a transformation rule demanded in lemma 4.3. Besides, for $b = 3/2$ we deal with $\sigma < 0$ and care is necessary in the sequel computations if logarithms or roots of σ are involved.

Corollary 4.4. Suppose $b = 3/2$ and $a = 0$. If

$$3\kappa\sqrt{\frac{\theta}{\gamma}} > 1. \tag{4.2}$$

then

$$\begin{aligned}
I(r_\tau : v_0) &\leq \frac{1}{12\sqrt{\alpha'} - 4\beta'} \int_{-\infty}^0 \frac{1}{\sigma^2} d\rho(\sigma) \\
\rho(\sigma_0) &= \frac{1}{Z} (-\sigma_0)^3 e^{-\alpha'(\sigma_0^2 - \beta'/\alpha')^2} \\
\rho(v_0) &= \frac{1}{Z'} \frac{1}{v_0^3} e^{-\alpha(1/v_0 - 1/\theta)^2} \\
\mathcal{I}(v_\tau, v_0) &= h(v_0) - \frac{1}{2} \log(2e\pi\kappa^2\tau) - \frac{3}{2} \int \log v_0 d\rho(v_0)
\end{aligned}$$

with

$$\begin{aligned}
\alpha' &= \frac{\gamma\kappa^2\theta}{64} \\
\beta' &= \frac{\gamma}{8}. \\
\alpha &= \frac{\gamma\theta}{\kappa^2}
\end{aligned}$$

Proof.

$$\begin{aligned}
V'(\sigma) &= \sqrt{2} \left\{ -\frac{\gamma}{\kappa} \left(\theta \left(-\frac{\sigma\kappa}{\sqrt{8}} \right)^3 - \left(-\frac{\sigma\kappa}{\sqrt{8}} \right) \right) - \frac{3}{\sqrt{2}} \frac{1}{\sigma} \right\} \\
&= \frac{\gamma\sqrt{2}}{\kappa} \left(\theta \left(\frac{\sigma\kappa}{\sqrt{8}} \right)^3 - \frac{\sigma\kappa}{\sqrt{8}} \right) - \frac{3}{\sigma} \\
&= \frac{\gamma}{2} \left(\frac{\theta\kappa^2}{8} \sigma^3 - \sigma \right) - \frac{3}{\sigma} \\
&= 4(\alpha'\sigma^3 - \beta'\sigma) - \frac{3}{\sigma}
\end{aligned}$$

for

$$\begin{aligned}
\alpha' &= \frac{\gamma\kappa^2\theta}{64} \\
\beta' &= \frac{\gamma}{8}.
\end{aligned}$$

This implies

$$\begin{aligned}
V''(\sigma) &= 4(3\alpha'\sigma^2 - \beta') + \frac{3}{\sigma^2} \\
V(\sigma) &= \alpha'\sigma^4 - 2\beta'\sigma^2 - 3\log(-\sigma) = \alpha' \left(\sigma^2 - \frac{\beta'}{\alpha'} \right)^2 - \frac{\beta'^2}{\alpha'} - 3\log(-\sigma)
\end{aligned}$$

One can read of that e^{-V} is integrable and the Gibbs distribution Eq. (2.10) reads in this case

$$\rho(\sigma) = \frac{1}{Z} (-\sigma)^3 e^{-\alpha'(\sigma^2 - \beta'/\alpha')^2}.$$

Z needs to be computed numerically. We check Eq. (2.11) by computing the third derivative.

$$0 = V'''(\sigma') = 24\alpha'\sigma' - \frac{6}{\sigma'^3} \Leftrightarrow \sigma'^2 = \sqrt{\frac{1}{4\alpha'}}$$

which yields

$$\lambda \equiv V''(\sigma') = 4(3\alpha'\sigma'^2 - \beta') + \frac{3}{\sigma'^2} = 6\sqrt{\alpha'} - 4\beta' + 6\sqrt{\alpha'} = 12\sqrt{\alpha'} - 4\beta' \quad (4.3)$$

which is greater zero iff $3\sqrt{\alpha'} > \beta'$ that is

$$3\kappa\sqrt{\frac{\theta}{\gamma}} > 1.$$

If this condition holds true the logarithmic Sobolev inequality for the stationary distribution holds with λ as in Eq. (4.3) and we can apply theorem 3.5. For the same reasons as in lemma 4.3 we have $I(r_\tau : v_0) = I(r_\tau : \sigma_0)$ and by an analogous computation as in the proof of lemma 4.3 we obtain

$$I(r_\tau : \sigma_0) \leq \frac{1}{12\sqrt{\alpha'} - 4\beta'} \int \frac{1}{\sigma_0^2} d\rho(\sigma_0).$$

We compute the stationary for the variance v itself. According to lemma 4.2, we have

$$\sigma^2 = \frac{8}{v\kappa^2}$$

and therefore

$$2\sigma g'(v) = -\frac{8}{v^2\kappa^2}.$$

This yields the stationary distribution for v

$$\begin{aligned} \rho(v) &= \frac{1}{Z'} (-\sigma)^3 e^{-\alpha' \left(\frac{8}{v\kappa^2} - \frac{\beta'}{\alpha'} \right)^2} \frac{-1}{\sigma v^2} \\ &= \frac{1}{Z'} \frac{1}{v^3} e^{-\alpha(1/v - 1/\theta)^2} \end{aligned}$$

for a constant Z' and

$$\alpha = \alpha' \frac{64}{\kappa^4} = \frac{\gamma\theta}{\kappa^2}$$

□

Corollary 4.5. *Suppose $b = 3/2$ and $a = 1$. Then*

$$\begin{aligned} I(r_\tau : v_0) &\leq \frac{1}{2(\alpha - 1)} \\ \rho(v_0) &= \frac{\beta^\alpha}{\Gamma(\alpha)} v_0^{-\alpha-1} e^{-\beta/v_0} \\ \mathcal{I}(v_\tau, v_0) &= \alpha + \frac{1}{2} \log \left(\frac{\Gamma(\alpha)^2}{\beta} \right) - \left(\alpha - \frac{1}{2} \right) \Psi(\alpha) - \frac{1}{2} \log(2e\pi\kappa^2\tau) \end{aligned}$$

where Γ is the Gamma function, Ψ the Digamma function and

$$\begin{aligned} \alpha &= \frac{2\gamma}{\kappa^2} + 2 \\ \beta &= \frac{2\gamma\theta}{\kappa^2}. \end{aligned}$$

Proof.

$$\begin{aligned}
V'(\sigma) &= \sqrt{2} \left\{ -\frac{\gamma}{\kappa} \left(\theta \left(-\frac{\sigma\kappa}{\sqrt{8}} \right) - \left(-\frac{\sigma\kappa}{\sqrt{8}} \right)^{-1} \right) - \frac{3}{\sqrt{2}} \frac{1}{\sigma} \right\} \\
&= \frac{\gamma}{\kappa} \left(\theta \frac{\sigma\kappa}{2} - \frac{4}{\sigma\kappa} \right) - \frac{3}{\sigma} \\
&= 2\beta'\sigma - \frac{(2\alpha-1)}{\sigma}
\end{aligned}$$

with

$$\begin{aligned}
\alpha &= \frac{2\gamma}{\kappa^2} + 2 \\
\beta' &= \frac{\gamma\theta}{4}.
\end{aligned}$$

This implies

$$\begin{aligned}
V''(\sigma) &= 2\beta' + \frac{(2\alpha-1)}{\sigma^2} > 2\beta' \equiv \lambda \quad \text{for all } \sigma \\
V(\sigma) &= \beta'\sigma^2 - (2\alpha-1)\log(-\sigma)
\end{aligned}$$

and therefore the stationary distribution ρ fulfils the logarithmic Sobolev inequality Eq. (2.12) and reads

$$\rho(\sigma) = \frac{1}{Z} (-\sigma)^{2\alpha-1} e^{-\beta'\sigma^2}.$$

Hence, σ^2 is Gamma distributed with shape α and rate β' and we recognize the partition function

$$Z = \frac{\Gamma(\alpha)}{2\beta'^\alpha}.$$

with the Gamma function Γ . Lemma 4.3 proves the inequality for the mutual information $I(r_\tau : v_0)$. Finally, we have

$$\sigma^2 = g(v)^2 = \frac{8}{\kappa^2} \frac{1}{v}$$

and therefore the variance v has the stationary distribution

$$\begin{aligned}
\rho(v) &= \frac{2\beta'^\alpha}{\Gamma(\alpha)} \left(\frac{8}{\kappa^2} \frac{1}{v} \right)^{\alpha-1/2} e^{-\beta' \frac{8}{\kappa^2} \frac{1}{v}} \frac{\sqrt{2}}{\kappa} \frac{1}{v^{3/2}} \\
&= \frac{2 \left(\beta' \frac{8}{\kappa^2} \right)^\alpha}{\Gamma(\alpha)} v^{-\alpha+1/2} v^{-3/2} e^{-\beta' \frac{8}{\kappa^2} \frac{1}{v}} \\
&= \frac{\beta'^\alpha}{\Gamma(\alpha)} v^{-\alpha-1} e^{-\beta/v},
\end{aligned}$$

that is, an inverse Gamma distribution with rate α and shape

$$\beta = \frac{8\beta'}{\kappa^2} = \frac{2\gamma\theta}{\kappa^2}$$

and we get

$$\begin{aligned}
h(v) &= \alpha + \log(\beta\Gamma(\alpha)) - (1+\alpha)\Psi(\alpha) \\
\int \log v d\rho(v) &= \log \beta - \Psi(\alpha)
\end{aligned}$$

where Ψ denotes the Digamma function. Lemma 4.1 yields the expression for the proxy $\mathcal{I}(v_\tau, v_0)$. \square

Corollary 4.6. *Suppose $b = 1$ and $a = 0$. Then*

$$\begin{aligned} I(r_\tau : v_0) &\leq \frac{1}{2(\alpha - 1)} \\ \rho(v_0) &= \frac{\beta^\alpha}{\Gamma(\alpha)} v_0^{-\alpha-1} e^{-\beta/v_0} \\ \mathcal{I}(v_\tau, v_0) &= \alpha(1 - \Psi(\alpha)) + \log \Gamma(a) - \frac{1}{2} \log(2e\pi\kappa^2\tau) \end{aligned}$$

with the Gamma function Γ , Digamma function Ψ and

$$\begin{aligned} \alpha &= \frac{2\gamma}{\kappa^2} + 1 \\ \beta &= \frac{2\gamma\theta}{\kappa^2} \end{aligned}$$

Proof.

$$\begin{aligned} V'(\sigma) &= \sqrt{2} \left\{ -\frac{\gamma}{\kappa} \left(\theta e^{\frac{-\kappa}{\sqrt{2}}\sigma} - 1 \right) + \frac{\kappa}{2} \right\} \\ &= -\beta \frac{\kappa}{\sqrt{2}} e^{\frac{-\kappa}{\sqrt{2}}\sigma} + \frac{\kappa}{\sqrt{2}} \left(\frac{2\gamma}{\kappa^2} + 1 \right) \\ &= -\beta \frac{\kappa}{\sqrt{2}} e^{\frac{-\kappa}{\sqrt{2}}\sigma} + \alpha \frac{\kappa}{\sqrt{2}} \end{aligned}$$

with

$$\begin{aligned} \alpha &= \frac{2\gamma}{\kappa^2} + 1 \\ \beta &= \frac{2\gamma\theta}{\kappa^2} \end{aligned}$$

and therefore

$$\begin{aligned} V''(\sigma) &= \beta \frac{\kappa^2}{2} e^{\frac{-\kappa}{\sqrt{2}}\sigma} \\ V(\sigma) &= \beta e^{\frac{-\kappa}{\sqrt{2}}\sigma} + \alpha \frac{\kappa}{\sqrt{2}} \sigma. \end{aligned}$$

One can read off that V'' is not uniformly bounded away from zero, and e^{-V} is integrable. If we define

$$\sigma' = \frac{1}{\sqrt{v}} = e^{-\frac{\kappa}{\sqrt{8}}\sigma}$$

then the stationary distribution ρ Eq. (2.10) for σ' reads

$$\begin{aligned} \rho(\sigma') &= \frac{1}{Z} e^{-V(\sigma)} e^{\frac{\kappa}{\sqrt{8}}\sigma} \\ &= \frac{1}{Z} e^{-\beta\sigma'^2} e^{-\alpha \frac{\kappa}{\sqrt{2}}\sigma} e^{\frac{\kappa}{\sqrt{8}}\sigma} \\ &= \frac{1}{Z} e^{-\frac{\kappa}{\sqrt{8}}\sigma(2\alpha-1)} e^{-\beta\sigma'^2} \\ &= \frac{1}{Z} \sigma'^{2\alpha-1} e^{-\beta\sigma'^2}. \end{aligned}$$

That is σ'^2 is Gamma distributed with shape α and rate β . Lemma 4.3 yields

$$Z = \frac{\Gamma(\alpha)}{2\beta^\alpha}$$

$$I(r_\tau : \sigma'_0) \leq \frac{1}{2(\alpha - 1)}$$

As in the case $b = 3/2, a = 1$ one checks that the variance v of the Stock process is inverse Gamma distributed with shape α and rate β . \square

Corollary 4.7. *Suppose $b = a = 1$ and*

$$\frac{2\gamma\theta}{\kappa^2} - 1 > 1.$$

Then

$$I(r_\tau : v_0) \leq \frac{1}{2(\alpha - 1)}$$

$$\rho(v_0) = \frac{\beta^\alpha}{\Gamma(\alpha)} v_0^{\alpha-1} e^{-\beta v_0}$$

$$\mathcal{I}(v_\tau, v_0) = \alpha(1 - \Psi(\alpha)) + \log \Gamma(\alpha) - \frac{1}{2} \log(2e\pi\kappa^2\tau)$$

with the Gamma function Γ , Digamma function Ψ and

$$\alpha = \frac{2\gamma\theta}{\kappa^2} - 1$$

$$\beta = \frac{2\gamma}{\kappa^2}$$

Proof.

$$V'(\sigma) = \sqrt{2} \left\{ -\frac{\gamma}{\kappa} \left(\theta - e^{\frac{\kappa}{\sqrt{2}}\sigma} \right) + \frac{\kappa}{2} \right\}$$

$$= \frac{2\gamma}{\kappa^2} \frac{\kappa}{\sqrt{2}} e^{\frac{\kappa}{\sqrt{2}}\sigma} + \frac{\kappa}{\sqrt{2}} \left(1 - \frac{2\gamma\theta}{\kappa^2} \right)$$

$$= \beta \frac{\kappa}{\sqrt{2}} e^{\frac{\kappa}{\sqrt{2}}\sigma} - \alpha \frac{\kappa}{\sqrt{2}}$$

with

$$\alpha = \frac{2\gamma\theta}{\kappa^2} - 1$$

$$\beta = \frac{2\gamma}{\kappa^2}$$

and therefore

$$V''(\sigma) = \beta \frac{\kappa^2}{2} e^{\frac{\kappa}{\sqrt{2}}\sigma}$$

$$V(\sigma) = \beta e^{\frac{\kappa}{\sqrt{2}}\sigma} - \alpha \frac{\kappa}{\sqrt{2}} \sigma.$$

V is not bounded from below by a positive real and e^{-V} is only integrable if $\alpha > 0$. In this case we introduce

$$\sigma' = \sqrt{v} = e^{\frac{\kappa}{\sqrt{8}}\sigma}$$

and the stationary distribution w.r.t. σ' reads

$$\begin{aligned}\rho(\sigma') &= \frac{1}{Z} e^{-V(\sigma)} e^{\frac{\kappa}{\sqrt{8}}\sigma} \\ &= \frac{1}{Z} e^{-\beta\sigma'^2} e^{-\alpha\frac{\kappa}{\sqrt{2}}\sigma} e^{\frac{\kappa}{\sqrt{8}}\sigma} \\ &= \frac{1}{Z} e^{-\frac{\kappa}{\sqrt{8}}\sigma(2\alpha-1)} e^{-\beta\sigma'^2} \\ &= \frac{1}{Z} \sigma'^{2\alpha-1} e^{-\beta\sigma'^2}.\end{aligned}$$

That is, the variance $v = \sigma'^2$ is Gamma distributed with shape α and rate β . The origin is reflecting iff $\alpha > 1$. For a Gamma distributed variance v with shape α and rate β we obtain

$$\begin{aligned}h(v) &= \alpha - \log \beta + \log \Gamma(\alpha) + (1 - \alpha)\Psi(\alpha) \\ \int \log v d\rho(v) &= \Psi(\alpha) - \log \beta\end{aligned}$$

The expression for the proxy $\mathcal{I}(v_\tau, v_0)$ then follows from lemma 4.1. □

Corollary 4.8. *Suppose $b = 1/2$, $a = 0$, and*

$$\frac{2\gamma\theta}{\kappa^2} > 1.$$

Then

$$\begin{aligned}I(r_\tau : v_0) &\leq \frac{1}{2(\alpha - 1)} \\ \rho(v_0) &= \frac{\beta^\alpha}{\Gamma(\alpha)} v_0^{\alpha-1} e^{-\beta v_0} \\ \mathcal{I}(v_\tau, v_0) &= \alpha - \frac{1}{2} \log \beta + \log \Gamma(\alpha) + \left(\frac{1}{2} - \alpha\right) \Psi(\alpha) - \frac{1}{2} \log(2e\pi\kappa^2\tau)\end{aligned}$$

with the Gamma function Γ , Digamma function Ψ and

$$\begin{aligned}\alpha &= \frac{2\gamma\theta}{\kappa^2} \\ \beta &= \frac{2\gamma}{\kappa^2}\end{aligned}$$

Proof.

$$\begin{aligned}V'(\sigma) &= \sqrt{2} \left\{ -\frac{\gamma}{\kappa} \left(\theta \left(\frac{\sigma\kappa}{\sqrt{8}} \right)^{-1} - \left(\frac{\sigma\kappa}{\sqrt{8}} \right) \right) + \frac{1}{\sqrt{2}}\sigma \right\} \\ &= \frac{\gamma}{2}\sigma - \frac{4\theta\gamma}{\kappa^2} \frac{1}{\sigma} + \frac{1}{\sigma} \\ &= 2\beta'\sigma + (1 - 2\alpha)\frac{1}{\sigma}\end{aligned}$$

with

$$\alpha = \frac{2\gamma\theta}{\kappa^2}$$

$$\beta' = \frac{\gamma}{4}.$$

Hence,

$$V''(\sigma) = 2\beta' - (1 - 2\alpha)\frac{1}{\sigma^2}$$

$$V(\sigma) = \beta'\sigma^2 + (1 - 2\alpha)\log \sigma.$$

e^{-V} is integrable iff $\alpha > 1/2$. Then σ^2 is Gamma distributed with shape α and rate β' , i.e.,

$$\rho(\sigma) = \frac{2\beta'^\alpha}{\Gamma(\alpha)}\sigma^{2\alpha-1}e^{-\beta'\sigma^2}$$

and the origin is reflecting iff $\alpha > 1$. The variance v is Gamma distributed with rate α and shape

$$\beta = \frac{8}{\kappa^2}\beta' = \frac{2\gamma}{\kappa^2}.$$

Everything else follows from lemma 4.3 and lemma 4.1. □

Corollary 4.9. *Suppose $b = 1/2$ and $a = 1$. Then, the Fokker-Planck equation*

$$\partial_t p = -\partial_v(\gamma v(\theta - v)p) + \frac{1}{2}\partial_{vv}(\kappa^2 vp)$$

has no stationary solution.

Proof.

$$\begin{aligned} V'(\sigma) &= \sqrt{2} \left\{ -\frac{\gamma}{\kappa} \left(\theta \left(\frac{\sigma\kappa}{\sqrt{8}} \right) - \left(\frac{\sigma\kappa}{\sqrt{8}} \right)^3 \right) + \frac{1}{\sqrt{2}} \frac{1}{\sigma} \right\} \\ &= -\frac{\gamma}{\kappa} \left(\theta \frac{\sigma\kappa}{2} - \left(\frac{\sigma\kappa}{\sqrt{8}} \right)^3 \right) + \frac{1}{\sigma} \\ &= -\frac{\gamma\theta}{2}\sigma + \frac{\gamma\kappa^2}{16}\sigma^3 + \frac{1}{\sigma} \end{aligned}$$

which yields

$$V(\sigma) = -\frac{\gamma\theta}{4}\sigma^2 + \frac{\gamma\kappa^2}{64}\sigma^4 + \log \sigma$$

and therefore

$$e^{-V(\sigma)} = \frac{1}{\sigma} e^{-\frac{\gamma\kappa^2}{64}\sigma^4 + \frac{\gamma\theta}{4}\sigma^2}$$

which is not an integrable function on $(0, \infty)$. The coordinate change g in lemma 4.2 is a differentiable and strictly monotonic. Hence, there is a stationary solution ρ if and only if

$$\rho(\sigma) = \rho(g^{-1}(\sigma)) g'(g^{-1}(\sigma)),$$

with $\sigma = g(v)$, is a stationary solution of the gradient flow Eq. (2.9). Assume there is a stationary solution ρ of Eq. (2.9) with boundaries x_{\min} and x_{\max} . If x_{\max} is finite, then $V(x_{\max}) = \pm\infty$ – see chapter 5 in [22] for this point. The same holds true for x_{\min} . Thus, the only possible boundaries are $x_{\min} = 0$ and $x_{\max} = \infty$. Hence, due to lemma 2.2, the Gibbs distribution Eq. (2.10) is defined on $(0, \infty)$ and e^{-V} is an integrable function on $(0, \infty)$ – a contradiction. □

Remark 4.2. As far as the authors know, this is the first time that non-existence of a stationary solution for the case $b = 1/2$ and $a = 1$ has been proven. See also [8] for a short discussion of this issue.

We have been assembled all ingredients for actually computing the mutual informations Eq. (3.3) for the five models which have (possibly) a stationary distribution. The parameters for the models were fitted [8] on Options of the S&P 500

a	b	γ	θ	κ	λ	ρ
0	1/2	3.1146	0.0523	0.5826	$9.00e-5$	-0.6520
		$\pm 1.50e-3$	$\pm 3.13e-5$	$\pm 1.27e-4$	$\pm 2.16e-3$	$\pm 2.41e-4$
0	1	2.4730	0.0272	1.1884	$4.37e-3$	-0.7116
		$\pm 1.12e-3$	$\pm 6.84e-6$	$\pm 3.20e-4$	$\pm 1.17e-3$	$\pm 2.06e-4$
1	1	64.4378	0.0367	1.1214	$1.93e-3$	-0.6749
		$\pm 3.95e-2$	$\pm 1.91e-5$	$\pm 4.36e-4$	$\pm 4.94e-2$	$\pm 2.50e-4$
0	3/2	1.5384	0.0336	7.9501	$7.91e-4$	-0.7169
		$\pm 2.22e-3$	$\pm 3.65e-5$	$\pm 2.94e-3$	$\pm 2.15e-3$	$\pm 2.10e-4$
1	3/2	50.9140	0.0388	6.2593	$3.36e-4$	-0.6854
		$\pm 4.08e-2$	$\pm 2.86e-5$	$\pm 2.41e-3$	$\pm 5.13e-2$	$\pm 2.35e-4$

Since the standard errors are comparably small, we shall ignore them in the sequel and compute only information values w.r.t. the means of the parameters. Furthermore, the risk premium λ cannot significantly distinguished from 0. Hence, we omit it. Recall, we are interested in the mutual informations

$$I(\sqrt{v_t\tau} : \mathbf{r}_{t-n\tau}^t)$$

$$I(r_t : \mathbf{r}_{t-n\tau}^0)$$

between the volatility $\sqrt{v_t\tau}$ and previous returns

$$\mathbf{r}_{t-n\tau}^t = r_{t-n\tau}, r_{t-(n-1)\tau}, \dots, r_t,$$

and previous returns and their subsequent return, respectively. In section 3.1 we derived in proposition 3.2 and 3.3 upper bounds U_1 for the mutual information $I(\sqrt{v_t\tau} : \mathbf{r}_{t-n\tau}^t)$ and U_2 for the mutual information $I(r_t : \mathbf{r}_{t-n\tau}^0)$, respectively. We computed in theorem 3.4 and 3.5 these upper bounds in terms of the parameters of the SDE Eq. (2.6) and the stationary distribution of the volatility process. Since the parameters γ , θ and κ from [8] are those for an annual time resolution, we have to choose

$$\tau = \frac{1}{252} \quad (4.4)$$

because we are dealing with daily returns. The previous corollaries provide explicit formulae for the upper bounds U_1 and U_2 which can be computed with ease in the case the stationary distribution exists. One checks that for the parameter values in the previous table, there is a stationary distribution in the cases $a = 0; b = 3/2$, $a = 1; b = 1$ but not for the Heston model $a = 0$ and $b = 1/2$. We list the values for the upper bounds U_1 and U_2 for the various models including the values G_r and $G_{\sqrt{v}}$ in Eq. (3.4) at the end of section 3.1.

a	b	α	β	U_2	U_1	G_r	$G_{\sqrt{v}}$
0	1/2	0.9598	18.35	—	—	—	—
0	1	4.502	0.09526	0.1428	2.230	12.67	7.318
1	1	2.761	102.5	0.2839	2.506	16.14	7.552
0	3/2	$8.511e-4$	—	0.1167	1.910	11.53	7.343
1	3/2	4.599	0.1008	0.1389	2.190	11.54	7.323

One sees, past returns never accumulate enough information about volatility. Dependencies between past and future returns are weak in all cases. Even though we have already observed in remark 3.2 that decreasing the time τ subsequent returns are read off does not affect the upper bound U_2 for the mutual information between past and future returns, the upper bound U_1 for the mutual information between past returns and the current volatility v_t grows by increasing the observation frequency for the returns. Hence, instead of asserting that daily returns do not yield enough information about their volatility the subsequent table lists necessary returns per annum, i.e. $1/\tau$, to obtain enough information from returns about volatility.

a, b	0, 1	1, 1	0, 3/2	1, 3/2
$1/\tau$	6.62 e6	6.08 e6	1.32 e7	7.24 e6

In all cases we need return data which is quoted every two or three seconds.

4.2 Exponential Ornstein-Uhlenbeck process

4.2.1 One-Factor Model

Another popular mean-reverting model is [21]

$$\begin{aligned} dS_t &= S_t m e^{v_t} dW_t^0 \\ dv_t &= -\gamma v_t dt + \kappa dW_t^1. \end{aligned}$$

The process v_t is an Ornstein-Uhlenbeck process. The solution of the corresponding Fokker-Planck equation with initial condition v_0 is

$$p(v_\tau | v_0) = \sqrt{\frac{2\gamma}{2\pi\kappa^2(1 - e^{-2\gamma\tau})}} e^{-\frac{\gamma}{\kappa^2} \frac{(v_\tau - v_0 e^{-\gamma\tau})^2}{1 - e^{-2\gamma\tau}}}, \quad (4.5)$$

that is, a normal distribution with mean $v_0 e^{-\gamma\tau}$ and variance

$$\frac{\kappa^2}{2\gamma} (1 - e^{-2\gamma\tau}).$$

The Ornstein-Uhlenbeck process has the stationary distribution

$$\rho(v) = \sqrt{\frac{\gamma}{\pi\kappa^2}} e^{-\gamma v^2 / \kappa^2},$$

that is, a normal distribution with mean 0 and variance

$$\frac{\kappa^2}{2\gamma}. \quad (4.6)$$

Corollary 4.10. *We have*

$$\begin{aligned} I(v_\tau : v_0) &= -\frac{1}{2} \log(1 - e^{-2\gamma\tau}) \\ I(r_\tau : v_0) &\leq \frac{2\gamma}{\kappa^2} \end{aligned}$$

Proof. Since we know the analytical expression Eq. (4.5) for the transition probability, we can compute the mutual information $I(v_\tau : v_0)$ precisely without the proxy 3.1 by means of example 2.1.

$$\begin{aligned} I(v_\tau : v_0) &= h(v_\tau) - h(v_\tau | v_0) \\ &= \frac{1}{2} \log \left(\pi e \frac{\kappa^2}{\gamma} \right) - \frac{1}{2} \left(\pi e (1 - e^{-2\gamma\tau}) \frac{\kappa^2}{\gamma} \right) \\ &= -\frac{1}{2} \log (1 - e^{-2\gamma\tau}) \end{aligned}$$

Furthermore, if we set $\rho(v) = e^{-V(v)}$ we obtain

$$V(v) = \frac{\gamma}{\kappa^2} v^2 + Z$$

for some constant Z and therefore

$$V''(v) = \frac{2\gamma}{\kappa^2}.$$

Hence, the stationary distribution fulfils the logarithmic Sobolev inequality with

$$\lambda = \frac{2\gamma}{\kappa^2},$$

and we obtain from theorem 3.5

$$\begin{aligned} I(r_\tau : v_0) &\leq \frac{1}{2\lambda} \int 2 \left| \frac{d}{dv} \log m e^v \right|^2 d\rho(v) \\ &= \frac{\kappa^2}{2\gamma}. \end{aligned}$$

□

From [21] we obtain for

$$\tau = 1 \text{ day}$$

the mean of the parameter values

$$\gamma = 1.82 e - 3 \text{ days}^{-1} \quad \kappa^2 = 1.4 e - 2 \text{ days}^{-1} \quad \rho = -0.4 \quad m = 1.5 e - 3 \text{ days}^{-1/2}$$

where $\rho dt = dW_t^0 dW_t^1$. Corollary 4.10 and theorems 3.4 and 3.5, respectively, yield

$$I(m e^{v_\tau} : \mathbf{r}_{\tau-n\tau}^\tau) \leq I(v_\tau : v_0) + I(r_\tau : v_\tau | v_0) = -\frac{1}{2} \log (1 - e^{-2\gamma\tau}) - \frac{1}{2} \log(1 - \rho^2) = 2.9$$

and

$$I(r_\tau : \mathbf{r}_{\tau-n\tau}^\tau) \leq I(r_\tau : v_0) \leq \frac{2\gamma}{\kappa^2} = 3.85.$$

The bound on the mutual information $I(r_\tau : v_0)$ is very weak. Hence we compute the mutual information numerically. We obtain from example 2.1

$$\begin{aligned} I(r_\tau : v_0) &= h(r_\tau) - h(r_\tau | v_0) \\ &= h(r_\tau) + \int p(r_\tau | v_0) \log p(r_\tau | v_0) dr_\tau d\rho(v_0) \\ &= h(r_\tau) - \frac{1}{2} \int \log (2\pi e m^2 e^{2v_0}) d\rho(v_0) \\ &= h(r_\tau) - \frac{1}{2} \log(2\pi e m^2) \\ &= 0.86 \end{aligned} \tag{4.7}$$

where the entropy $h(r_\tau)$ was computed numerically, recalling that the distribution of r_τ is

$$p(r_\tau) = \int p(r_\tau|v_0)\rho(v_0)dv_0 = \int \frac{1}{\sqrt{2\pi m^2 e^{2v_0}}} e^{-\frac{r_\tau^2}{2m^2 e^{2v_0}}} \sqrt{\frac{\gamma}{\pi \kappa^2}} e^{-\gamma v_0^2/\kappa^2} dv_0. \quad (4.8)$$

Since we deal with daily returns, we compute the necessary information G_r and G_{me^v} which we need for quoting the returns and the volatility up to 0.01 percent precision as in Eq. (3.5) and obtain

$$G_r = \log\left(\frac{\sigma_f}{\sigma_M}\right) = \log\left(\frac{me^{\frac{\kappa^2}{4\gamma}}}{\sigma_M}\right) = 6.0$$

and

$$\begin{aligned} G_{me^v} &= h(v) + \int \log(me^v) d\rho(v) + \frac{1}{2} \log(252) - \frac{1}{2} \log(2\pi e \sigma_M^2) \\ &= \frac{1}{2} \log\left(2\pi e \frac{\kappa^2}{2\gamma}\right) + \log(m) + \frac{1}{2} \log(252) - \frac{1}{2} \log(2\pi e \sigma_M^2) \\ &= \log\left(\frac{\kappa}{\sqrt{2\gamma}\sigma_M}\right) + \log(m) + \frac{1}{2} \log(252) = 7.5 \end{aligned}$$

with $\sigma_M = \frac{1}{4}10^{-4}$. We summarize our results in the subsequent table. As previously, U_1 denotes the upper bound for the mutual information $I(me^{v_\tau} : \mathbf{r}_{\tau-n\tau}^\tau)$ and U_2 the one for the mutual information $I(r_\tau : \mathbf{r}_{\tau-n\tau}^\tau)$.

U_1	U_2	G_r	G_{me^v}
2.9	0.86	6.0	7.5

As for the previous models, estimating volatility in the Exponential Ornstein-Uhlenbeck setting from stock data is hardly possible. Again returns are only weakly dependent in this model. To obtain the required precision for the volatility we would need at least 9995 returns a day, i.e., we need nearly secondly quoted returns.

4.2.2 Two-Factor Model

As we show now, the situation does not improve in case of multi-factor models. In [1] a two-factor version of the above model, i.e.

$$\begin{aligned} dS_t &= S_t me^{v_{1,t}+v_{2,t}} dW_t^0 \\ dv_{1,t} &= -\gamma_1 v_{1,t} dt + \kappa_1 dW_t^1 \\ dv_{2,t} &= -\gamma_2 v_{2,t} dt + \kappa_2 dW_t^2 \end{aligned}$$

is defined and compared to the above one-factor model. In this paper, all Brownian motion are assumed to be independent which we adopt for simplicity as well.

Regarding the information computations two complications arise when considering multi-factor models:

- Approximating the information between $\mathbf{v}_t = (v_{1,t}, v_{2,t})$ and $\mathbf{r}_{t-n\tau}^t$ by

$$I(\mathbf{v}_t : \mathbf{r}_{t-n\tau}^t) \leq I(\mathbf{v}_t : \mathbf{v}_{t-\tau}) + I(\mathbf{v}_t : r_t | \mathbf{v}_{t-\tau})$$

as in Prop. 3.2 would overstate the actual information since the returns do not depend on $v_{1,t}$ and $v_{2,t}$ directly, but only via the sum $v_{1,t} + v_{2,t}$ resulting in a loss of information.

- While the two-dimensional process \mathbf{v}_t is Markovian, this is no longer the case for the volatility process $\sigma_t = m e^{v_{1,t} + v_{2,t}}$. Thus, we need to adapt the bound from Prop. 3.2 in order to obtain a tighter bound on $I(\sigma_t : \mathbf{r}_{t-n\tau}^t)$ which is our prime interest.

Fortunately, the exponential Ornstein-Uhlenbeck model is fully tractable and even though $w_t = v_{1,t} + v_{2,t}$ is not Markovian, it is still a Gaussian process with stationary variance $\sigma_w^2 = \frac{\kappa_1^2}{2\gamma_1} + \frac{\kappa_2^2}{2\gamma_2}$ and covariance $c_{w,n} = \frac{\kappa_1^2}{2\gamma_1} e^{-\gamma_1 n\tau} + \frac{\kappa_2^2}{2\gamma_2} e^{-\gamma_2 n\tau}$ between w_t and $w_{t+n\tau}$. In particular, we can compute the conditional distribution of w_t conditioned on $\mathbf{w}_{t-n\tau}^{t-\tau}$ and thus the mutual information $I(w_t : \mathbf{w}_{t-n\tau}^{t-\tau}) = I(\sigma_t : \boldsymbol{\sigma}_{t-n\tau}^{t-\tau})$.

Proposition 4.11.

$$I(\sigma_t : \mathbf{r}_{t-n\tau}^t) \leq I(w_t : \mathbf{w}_{t-n\tau}^{t-\tau})$$

Proof. We can bound the information as follows:

$$\begin{aligned} I(\sigma_t : \mathbf{r}_{t-n\tau}^t) &= I(w_t : \mathbf{r}_{t-n\tau}^t) \\ &\leq I(w_t : \mathbf{w}_{t-n\tau}^{t-\tau}, \mathbf{r}_{t-n\tau}^t) \\ &= I(w_t : \mathbf{w}_{t-n\tau}^{t-\tau}) + I(w_t : \mathbf{r}_{t-n\tau}^t | \mathbf{w}_{t-n\tau}^{t-\tau}) \\ &= I(w_t : \mathbf{w}_{t-n\tau}^{t-\tau}) + I(w_t : \mathbf{r}_{t-n\tau}^{t-\tau} | \mathbf{w}_{t-n\tau}^{t-\tau}) + I(w_t : r_t | \mathbf{w}_{t-n\tau}^{t-\tau}, \mathbf{r}_{t-n\tau}^{t-\tau}) \\ &= I(w_t : \mathbf{w}_{t-n\tau}^{t-\tau}) + I(w_t : r_t | \mathbf{w}_{t-n\tau}^{t-\tau}) \end{aligned}$$

where we have used that $w_t \perp \mathbf{r}_{t-n\tau}^{t-\tau} | \mathbf{w}_{t-n\tau}^{t-\tau}$, i.e. the past returns have no influence on w_t when the whole history $\mathbf{w}_{t-n\tau}^{t-\tau}$ is available. Note that as we assumed no leverage effect, i.e. $\rho_1 = \rho_2 = 0$, the last term vanishes. Thus, we are left with the bound based on the information structure of the Gaussian process. \square

Table 1 contains numeric values for the information $I(w_t : \mathbf{w}_{t-n\tau}^{t-\tau})$ using parameters from [1]:

$$\begin{aligned} m &= 2.32 e - 3 \text{ days}^{-\frac{1}{2}} & \gamma_1 &= 2.02 e - 2 \text{ days}^{-1} & \kappa_1^2 &= 4.13 e - 3 \text{ days}^{-1} \\ \gamma_2 &= 1.43 \text{ days}^{-1} & \kappa_2^2 &= 4.14 e - 2 \text{ days}^{-1} \end{aligned}$$

where all parameters have been converted to daily units based on their convention of 257 trading days per year. For illustration we have chosen the parameter values fitted on Canadian dollar exchange rates, since these give rise to the largest information values.

	history length n						
	1	2	3	4	5	10	100
$I(\sigma_t : \mathbf{r}_{t-n\tau}^t)$	0.778	0.819	0.835	0.842	0.845	0.847	0.847

Table 1: Upper bound on the mutual information (in nats) about the volatility when observing returns $\mathbf{r}_{t-n\tau}^t$.

Finally, we compute the mutual information $I(r_\tau : w_0)$ with $w_0 = v_{1,0} + v_{2,0}$ in the same manner as we computed the mutual information $I(r_\tau : v_0)$ for the single-factor mode in Eq. (4.7), namely numerically. We proceed by replacing v_0 by w_0 and the variance $\frac{\kappa^2}{2\gamma}$ by $\frac{\kappa_1^2}{2\gamma_1} + \frac{\kappa_2^2}{2\gamma_2}$ in formula Eq. (4.8) and obtain the value

$$I(r_\tau : w_0) = 0.093.$$

As before, we compute the required information G_r and G_{me^w} as in Eq. (3.5), again using parameters with daily units and obtain

$$G_r = \log\left(\frac{\sigma_f}{\sigma_M}\right) = \log\left(\frac{me^{\frac{\kappa_1^2}{4\gamma_1} + \frac{\kappa_2^2}{4\gamma_2}}}{\sigma_M}\right) = 4.6$$

and

$$\begin{aligned} G_{me^w} &= h(w) + \int \log(me^w) d\rho(w) + \frac{1}{2} \log(252) - \frac{1}{2} \log(2\pi e \sigma_M^2) \\ &= \frac{1}{2} \log\left(\frac{\frac{\kappa_1^2}{2\gamma_1} + \frac{\kappa_2^2}{2\gamma_2}}{\sigma_M^2}\right) + \log(m) + \frac{1}{2} \log(252) = 6.2 \end{aligned}$$

where $\sigma_M = \frac{1}{4}10^{-4}$.

Here, in contrast to the one-factor model the bound on the information increases when longer histories are observed. Nevertheless, the information values are much smaller, also in relation to the required information, than for the one-factor model considered above and the information essentially saturates after about 10 days. As explained in [1] the two-factor model improves the one-factor model by utilizing two very different time scales for the processes $v_{1,t}$ and $v_{2,t}$. In particular, one of these processes captures fast and transient changes of the volatility while the other models long-range dependencies. Especially the transient process diminishes the temporal dependence of the volatility process substantially. Thus, if the finding on the nature of two-factor models holds up in general, i.e. across asset classes, we would not be surprised if they provide even less information about the hidden volatility process than their, potentially misspecified, one-factor relatives.

5 Conclusions

We developed a general information theoretical frame to estimate in stochastic volatility models the uncertainty about the hidden volatility when it is inferred from stock data. This frame also allows to quantify the dependencies between subsequent returns in these models.

In single factor models, quoting volatility up to 0.01 percent needs in general at least secondly quoted return time series. Even then, only the upper bound, which we derived for the mutual information between past returns and present volatility, bridges the information gap computed in Eq. (3.4) not necessarily the actual mutual information itself. The situation does not improve when considering two-factor models. To the contrary, as we show in the case of a two-factor exponential Ornstein-Uhlenbeck model, even less information about the underlying volatility can be recovered from return data. The reason being that real volatility is a highly varying process, which is better captured by the more flexible two-factor model. Thus, instantaneous volatility cannot be estimated from much more than about ten days of return data. We note that these results also apply when predicting volatility, i.e. after about ten days the best prediction is based on the stationary distribution alone. This high intrinsic uncertainty of volatility estimates sheds doubt on the standard practice of comparing volatility models based on their forecast performance. To the least, much care is needed to obtain reliable and significant statements about the relative performance of different models. To our knowledge, this point has not been discussed in the literature let alone being studied in a rigorous and quantitative fashion. Here, we demonstrate that our information theoretic frame is ideally suited to address this

issue and derive precise data requirements, e.g. returns at second resolution, in order to obtain reliable volatility estimates.

Remarkable, from a technical perspective, is the upper bound we derived in theorem 3.5 for the mutual information between past returns and future ones by means of the logarithmic Sobolev inequality. Deriving such an inequality is interesting in its own right because it is in general pretty difficult to derive upper bounds for the mutual information because it captures all dependencies between two random variables and is therefore much harder to handle than, for instance, correlation. The use of the logarithmic Sobolev inequality demonstrates a nice interplay between recent developments in statistical physics where this inequality played a prominent role in estimating the convergence rate of perturbed thermodynamical systems towards their equilibrium and financial mathematics. As far as we know, the present paper is the first one which applies the logarithmic Sobolev inequality in finance. Apart from its mathematical charm the inequality yields an upper bound for the mutual information between past and future returns which does no longer depend on the time resolution τ and which is quite tight for the generic stochastic volatility models considered in section 4.1. Thus proving that despite the correlated volatility process returns are only weakly dependent in stochastic volatility models. This is not only consistent with the observation that returns are hard to predict, but might also motivate the use of simpler, single factor jump-diffusion models, which are analytically more tractable. We adopt in a forthcoming paper the information theoretical methods developed in the present one for estimating the information content of vanilla option prices about the underlying volatility. It turns out that volatility estimates from option prices are far superior over stock returns as more data is available: for the same volatility different options at different strikes and maturities are quoted, and furthermore the relationship between option prices and volatility is much more direct than between returns and volatility. Future research will be also devoted on the question to which extend *rough paths* stochastic volatility models, see for instance [3], which have recently caught attention, are in the information theoretical scope of the present paper.

6 Funding

Nils Bertschinger and Oliver Pfante thank Dr. h.c. Maucher for funding their positions.

References

- [1] S. Alizadeh, M. W. Brandt, and F. X. Diebold. Range-based estimation of stochastic volatility models. *Journal of Finance*, 57:1047–1091, 2002.
- [2] D. Bakry and M. Ledoux. Lévy-gromov’s isoperimetric inequality for an infinite-dimensional diffusion generator. *Invent. Math.*, 123(2):259–281, 1996.
- [3] C. Bayer, P. K. Friz, and J. Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2016.
- [4] F. Black and M. Scholes. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81(3):637, 1973.
- [5] T. Bollerslev, J. Litvinova, and G. Tauchen. Leverage and volatility feedback effects in high-frequency data. *Journal of Financial Econometrics*, 4(3):353–384, 2006.
- [6] J. P. Bouchaud, a. Matacz, and M. Potters. Leverage effect in financial markets: the retarded volatility model. *Physical review letters*, 87(22):228701, 2001.

- [7] J.-P. Bouchaud and M. Potters. Theory of Financial Risk and Derivative Pricing. *Theory of Financial Risk and Derivative Pricing*, page 379, 2003.
- [8] P. Christoffersen, K. Jacobs, and K. Mimouni. Models for s&p 500 dynamics: Evidence from realized volatility, daily returns, and option prices. *Review of Financial Studies*, 23(8), 2007.
- [9] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2 edition, July 2006.
- [10] Z. Ding, C. W. Granger, and R. F. Engle. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1):83–106, 1993.
- [11] W. Feller. Two Singular Diffusion Problems. *Annals of Mathematics*, 54(2):286–295, 1951.
- [12] J. Gatheral. Jump-diffusion models. *Encyclopedia of Quantitative Finance*, 2010.
- [13] Gross. Logarithmic sobolev inequalities. *Amer. J. Math.*, 97:1061–1083, 1975.
- [14] P. R. Hansen and A. Lunde. A forecast comparison of volatility models: does anything beat a garch(1,1)? *Journal of Applied Econometrics*, 20(7):873–889, 2005.
- [15] C. S. Jones. The dynamics of stochastic volatility: evidence from underlying and options markets. *Journal of Econometrics*, (116):181–224, 2003.
- [16] A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. *Phys. Rev. E*, (69), 2004.
- [17] B. LeBaron. Stochastic Volatility as a Simple Generator of apparent financial power laws and long memory. *Quantitative Finance*, 6(1):627–631, 2001.
- [18] A. W. Lo. Long Term Memory in Stock Market Prices. *Econometrica*, 59(5):1279–1313, 1991.
- [19] P. A. Markowich and C. Villani. On the trend to equilibrium for the fokker-planck equation : An interplay between physics and functional analysis.
- [20] J. F. Muzy, J. Delour, and E. Bacry. Modelling fluctuations of financial time series : from cascade process to stochastic volatility model. *Eur. Phys. J. B*, 17:537–548, 2000.
- [21] J. Perello, R. Sircar, and J. Masoliver. Option pricing under stochastic volatility: the exponential Ornstein-Uhlenbeck model. page 26, 2008.
- [22] H. Risken. *The Fokker-Planck Equation*. Springer, 1996.
- [23] W. Schoutens. *Lévy Processes in Finance: Pricing Financial Derivatives*. Wiley, 2003.
- [24] C. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 1948.
- [25] A. J. Stam. Some inequalities satisfied by the quantities of information of fisher and shannon. *Information and Control*, 2:101–112, 1959.